# Molecular Basis of Aortic Diseases

---

**Dissertation**

**zur**

**Erlangung der naturwissenschaftlichen Doktorwürde**
**(Dr. sc. nat.)**

**vorgelegt der**

**Mathematisch-naturwissenschaftlichen Fakultät**

**der**

**Universität Zürich**

**von**

Janine Meienberg

**von**

Neuheim ZG

**Promotionskomitee**

| | | |
|---|---|---|
| Prof. Dr. | Thierry Hennet | (verantwortliches Fakultätsmitglied) |
| PD Dr. | Gabor Matyas | (Leitung der Dissertation) |
| Prof. Dr. | Matthias Baumgartner | |
| Prof. Dr. | Sabina Gallati | |

**Zürich, 2015**

# Declaration

I declare that my thesis was composed by myself and the enclosed experimental work was performed on my own. Exceptions are explicitly stated in the text.
This dissertation has not been submitted for any other degree or professional qualification.

Janine Meienberg, Zurich 2015

# Preface

This work was carried out at the Center for Cardiovascular Genetics and Gene Diagnostics of the Foundation for People with Rare Diseases (2012-2015) and the Institute of Medical Molecular Genetics of the University of Zurich (2010-2011).

First of all, I would like to thank PD Dr. Gabor Matyas for taking me as a PhD student and enabling me to get insight into this hot field of research. It was for me a great opportunity and challenge to work with newest technologies and to perform exciting research of high clinical relevance. I really appreciate his confidence and patience as well as his support and the time he took for discussions.

A further thank goes to Prof. Dr. Thierry Hennet for serving as responsible faculty member despite his tight time schedule as well as to Prof. Dr. Matthias Baumgartner and Prof. Dr. Sabina Gallati for accompanying my PhD project as members of my PhD committee. I really appreciated their valuable inputs during the committee meetings.

A special thank goes to Dr. Benno Roethlisberger at the Division of Medical Genetics of the Center for Laboratory Medicine in Aarau for allowing me to perform microarray analyses and next-generation sequencing on the MiSeq in his lab as well as his team for technical support and helpful discussions.

I am grateful to Andrea Patrignani, Dr. Michal Okoniewski, Dr. Rémy Bruggmann, Dr. Steffen Zeisberger, and all the other collaboration partners for any help and constructive discussions.

I am grateful to the diagnostics team of the Center for Cardiovascular Genetics and Gene Diagnostics, including Nicole Amstutz, Eliane Arnold, Regina Perez, Philippe Reuge, and especially Caroline Henggeler, for their laboratory support and constructive discussions. In addtion, I would like to thank Stefan Widmer for establishing Sanger sequencing of the gene *B3GLCT* as well as Sunny Singh and Paulina Naef for their assistance in the breakpoint analyses of detected deletions and Andry Ehrhart for assistance in the analysis of NGS data. Moreover, I also thank all the current and previous members of the center for good incorporation and helpful discussions. It is really a pleasure to be a part of this team.

I am also grateful to all the foundations which gave the financial support needed to conduct this PhD project.

Last but not least, I would like to thank my parents, my sister, my brother and his family, as well as my friends for their support, understanding, and patience during this time.

# Table of Contents

# Zusammenfassung

Aortenkrankheiten (AD), welche Aneurysmen und Dissektionen der Aorta umfassen, sind aufgrund ihres erhöhten Risikos für Aortenrupturen mit hoher Morbidität und Mortalität assoziiert. AD können infolge von verschiedenen Risikofaktoren spontan oder in Zusammenhang mit einer genetisch bedingten (Bindegewebs-)Krankheit auftreten. Eine Vielzahl solch hereditärer Formen von AD wurde bereits beschrieben. Diese können, müssen aber nicht, mit zusätzlich betroffenen Organsystemen einhergehen und ein überlappendes klinisches Erscheinungsbild aufweisen. Die bislang gängigen genetischen Abklärungen mittels Sanger-Sequenzierung und der «*multiplex ligation-dependent probe amplification*» (MLPA) Methode sind durch die hohe Anzahl und Grösse der mit AD assoziierten Gene sowie durch das biologische bzw. wissenschaftliche Problem erschwert, dass bisher nicht alle mit AD assoziierten Gene bekannt sind. Deshalb war das Ziel dieser Doktorarbeit, mittels neuer Hochdurchsatz-Verfahren zum besseren Verständnis der molekularen Grundlagen von Aortenkrankheiten beizutragen.

Eine Kohorte von 65 AD-Patienten, bei welchen die bisherige genetische Abklärung keine Mutation aufgezeigt hat, wurde mit Microarrays (aCGH) untersucht, deren Proben für exonische Regionen angereichert sind, um auch Deletionen von einzelnen Exons erkennen zu können. Dabei wurden keine exonische Deletionen in bereits mit AD assoziierten Genen nachgewiesen, dafür aber bei 11 Patienten (11/65) in möglichen Kandidatengenen. Diese Gene, deren Rolle bei AD durch weitere Untersuchungen geklärt werden muss, sind direkt bei der Zusammensetzung der extrazellulären Matrix (z.B. *PCDHGB4* und *VWA3A*) oder im TGFβ-Signalweg (z.B. *FGFR2* und *B3GLCT*), der bei AD verändert ist, involviert. Da die Charakterisierung grosser Deletionen mit Long-Range PCR und Sanger-Sequenzierung arbeitsintensiv und zeitaufwendig ist, wurde die Hochdurchsatz-Sequenzierung (NGS) für die Bestimmung der Bruchpunkte angewendet. Es konnte gezeigt werden, dass hierzu sowohl Sequenziergeräte der zweiten, wie auch der dritten Generation eine effektive und effiziente Alternative darstellen.

Da aber die meisten AD-Fälle durch «*single nucleotide variants*» (SNVs) verursacht werden, wurde die Hochdurchsatz-Sequenzierung des ganzen Exoms (WES) sowie des ganzen Genoms (WGS) evaluiert. Unsere Daten weisen darauf hin, dass WGS die hohen Ansprüche von genetischen Abklärungen besser erfüllt als WES. Sie deckt nicht nur kodierende Exons, vor allem in GC-reichen Regionen, besser ab, sondern ist auch bei der Detektion von strukturellen Varianten (SV) und Abweichungen im nicht-kodierenden Bereich besser geeignet. WGS stellt im Vergleich zu Microarrays zudem eine effizientere Untersuchungsmethode dar, da sie gleichzeitig SNVs und SVs detektieren kann. Das Analysieren und Interpretieren der NGS/WGS-Daten bleibt aber eine Herausforderung und ist Gegenstand aktueller Forschung.

Diese Doktorarbeit gibt neue Einblicke in die Erforschung der molekularen Grundlagen von Aortenkrankheiten und eröffnet den Weg zur Erfassung von neuen (Kandidaten-)Genen sowie zur Entwicklung von neuen Therapien für AD.

# Summary

Aortic diseases (AD) including aortic aneurysms and dissections are associated with high morbidity and mortality due to an increased risk for aortic ruptures. AD can occur spontaneously due to several risk factors or in association with a genetic (connective tissue) disorder. So far, numerous heritable forms of AD have been described, which can be non-syndromic with no extra-cardiovascular features or syndromic involving multiple organ systems with overlapping clinical phenotypes. Traditional diagnostic testing using Sanger sequencing and multiplex ligation-dependent probe amplification (MLPA) is hampered by the high number and large size of genes associated with AD as well as by the biological or rather scientific problem that some of the genes mutated in AD are still unknown. Consequently, the aim of this thesis was to contribute to a better understanding of the molecular basis of aortic diseases by using novel high-throughput technologies.

A cohort of 65 AD patients with no mutation detected using previous standard genetic testing was screened by microarrays (aCGH) custom designed for exonic regions to enable the detection of deletions affecting single exons. No deletions in genes known to be associated with AD have been identified. However, in 11 patients (11/65) deletions in genes directly playing a role in the composition of the extracellular matrix (e.g. *PCDHGB4* and *VWA3A*) or in TGFβ signalling (e.g. *FGFR2* and *B3GLCT*), which is known to be involved in the pathogenesis of AD, have been detected. Further examinations are needed to determine the role of these genes in AD. As the characterization of such deletions using long-range PCR and Sanger sequencing is often laborious and time-consuming, next-generation sequencing (NGS) was applied for the identification of deletion breakpoints. It was shown that both second- and third-generation sequencing platforms provide an effective and efficient alternative for the characterization of deletion breakpoints.

Since most AD cases are caused by single nucleotide variants (SNVs) or small insertions and deletions, high-throughput sequencing approaches like the sequencing of the whole exome (WES), i.e. all known exons in the human genome, or the whole genome (WGS) were also evaluated. Our data suggest that WGS fulfils the needs of molecular diagnostics better than WES. It not only outperforms WES in sufficiently covering coding exons, especially GC-rich regions, but is also more powerful for the detection of structural variants (SVs) as well as sequence variants in non-coding regions. Furthermore, WGS represents a more efficient screening method than microarrays as it allows the simultaneous detection of both SNVs and SVs. However, the analysis and interpretation of NGS/WGS data remains challenging and is still subject to ongoing research.

This thesis provides novel insights into the assessment of the molecular basis of aortic diseases and opens the way for the identification of novel (candidate) genes and the development of novel therapies for AD.

# Abbreviations

| | |
|---|---|
| A | Adenine |
| AAA | Abdominal aortic aneurysm |
| ACE | Accuracy and Content Enhanced |
| ACEI | Angiotensin-converting enzyme inhibitor |
| aCGH | Array comparative genomic hybridization |
| AD | Aortic disease |
| ADAMTS-13 | A disintegrin and metalloproteinase with thrombospondin motifs 13 |
| ADAMTSL-1 | ADAMTS-like protein 1 |
| *ACTA2* | Actin, alpha 2, smooth muscle, aorta; MIM *102620 |
| *APC* | Adenomatous polyposis coli; MIM *611731 |
| ARB | Angiotensin II type 1 receptor blocker |
| *ARVCF* | Armadillo repeat gene deleted in velocardiofacial syndrome; MIM *602269 |
| ATR1 | Angiotensin II type 1 receptor |
| ATR2 | Angiotensin II type 2 receptor |
| ATS | Arterial tortuosity syndrome; MIM #208050 |
| *B3GLCT* | Beta 3-glucosyltransferase; MIM *610308 |
| bp | Base pairs |
| C | Cytosine |
| *CDKN2B* | Cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4); MIM *600431 |
| *CHD7* | Chromodomain helicase DNA binding protein 7; MIM *608892 |
| ChIP-seq | Chromatin immunoprecipitation followed by NGS |
| *CNTNAP2* | Contactin associated protein-like 2; MIM *604569 |
| CNV | Copy number variant |
| co-SMAD | Common SMAD |
| *COL3A1* | Collagen, type III, alpha 1; MIM *120180 |
| *COL6A5* | Collagen, type VI, alpha 5; MIM *611916 |
| *COL9A3* | Collagen, type IX, alpha 3; MIM *120270 |
| CRISPR/Cas | Clustered regularly interspaced short palindromic repeats/CRISPR-associated |
| DGV | Database of Genomic Variants |
| DNA | Deoxyribonucleic acid |
| EDS III | Ehlers-Danlos syndrome, hypermobility type; MIM %130020 |
| EDS IV | Ehlers-Danlos syndrome, vascular type; MIM #130050 |
| EDTA | Ethylenediaminetetraacetic acid |

| | |
|---|---|
| ECM | Extracellular matrix |
| ERK | Extracellular signal-regulated kinase |
| *FBN1* | Fibrillin 1; MIM *134797 |
| FASST2 | Fast Adaptive States Segmentation Technique 2 |
| *FGFR2* | Fibroblast growth factor receptor 2; MIM *176943 |
| FoSTeS | Fork Stalling and Template Switching |
| G | Guanine |
| gDNA | Genomic DNA |
| GLUT10 | Glucose transporter 10 |
| GWAS | Genome-wide association study |
| *HAS1* | Hyaluronan synthase 1; MIM *601463 |
| *HOXA1* | Homeobox A1; MIM *142955 |
| *HSPG2* | Heparan sulfate proteoglycan 2; MIM *142461 |
| INDEL | Small insertion and deletion |
| *ITGAE* | Integrin, alpha E (antigen CD103, human mucosal lymphocyte antigen 1; alpha polypeptide); MIM *604682 |
| JNK | c-Jun N-Terminal kinase |
| K | Thousand |
| kb | Kilobases |
| *KMT2D* | Lysine (K)-specific methyltransferase 2D; MIM *602113 |
| LAP | Latency-associated peptide |
| LDS | Loeys-Dietz syndrome; MIM #609192, #610168, #613795, #614816 |
| *LEFTY2* | Left-right determination factor 2; MIM +601877 |
| LLC | Large latent complex |
| lncRNA | Long non-coding RNA |
| LOH | Loss of heterozygosity |
| LR-PCR | Long-range PCR |
| LTBP | Latent TGFβ-binding protein |
| *LTBP3* | Latent transforming growth factor beta binding protein 3; MIM *602090 |
| *LTBP4* | Latent transforming growth factor beta binding protein 4; MIM *604710 |
| M | Million |
| MAGP-2 | Microfibril-associated glycoprotein-2 |
| MAPK | Mitogen-activated protein kinase |
| MAT | Methionine adenosyltransferase |
| *MAT2A* | Methionine adenosyltransferase II, alpha; MIM *601468 |
| Mb | Megabases |

| | |
|---|---|
| *MFAP5* | Microfibrillar associated protein 5; MIM *601103 |
| MFS | Marfan syndrome; MIM #154700 |
| MLCK | Myosin light chain kinase |
| MLPA | Multiplex ligation-dependent probe amplification |
| MMP | Matrix metalloproteinase |
| *MMP26* | Matrix metalloproteinase 26; MIM *605470 |
| *MYH11* | Myosin, heavy chain 11, smooth muscle; MIM *160745 |
| *MYLK* | Myosin light chain kinase; MIM *600922 |
| *MYO19* | Myosin XIX |
| NAHR | Non-allelic homologous recombination |
| *NDUFA6* | NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 6, 14kDa; MIM *602138 |
| NGS | Next-generation sequencing |
| NHEJ | Non-homologous end joining |
| NMD | Nonsense-mediated mRNA decay |
| *NOS1* | Nitric oxide synthase 1 (neuronal); MIM *163731 |
| *NOTCH1* | Notch 1; MIM *190198 |
| *NOTCH2* | Notch 2; MIM *600275 |
| *NOTCH3* | Notch 3; MIM *600276 |
| *NPHP3* | Nephronophthisis 3 (adolescent); MIM *608002 |
| *OR51L1* | olfactory receptor, family 51, subfamily L, member 1 |
| PCDHG | Protocadherin-gamma gene cluster; MIM #604968 |
| *PCDHGA8* | Protocadherin gamma, subfamily A, 8; MIM *606295 |
| *PCDHGA11* | Protocadherin gamma, subfamily A, 11; MIM *606298 |
| *PCDHGB4* | Protocadherin gamma, subfamily B, 4; MIM *603058 |
| *PCDHGB5* | Protocadherin gamma, subfamily B, 5; MIM *606302 |
| PCR | Polymerase chain reaction |
| *PKD1* | Polycystic kidney disease 1 (autosomal dominant); MIM *601313 |
| PKG-1 | Type I cGMP-dependent protein kinase |
| *PRKG1* | Protein kinase, cGMP-dependent, type I; MIM *176894 |
| R-SMAD | Receptor-regulated SMAD |
| *RAG1* | Recombination activating gene 1; MIM *179615 |
| *RAG2* | Recombination activating gene 2; MIM *179616 |
| RNA | Ribonucleic acid |
| ROCK | Rho-associated protein kinase |
| SGS | Shprintzen-Goldberg syndrome; MIM #182212 |

| | |
|---|---|
| *SKI* | SKI proto-oncogene; MIM *164780 |
| SLC | Small latent complex |
| *SLC2A10* | Solute carrier family 2 (facilitated glucose transporter), member 10; MIM *606145 |
| *SMAD3* | SMAD family member 3; MIM *603109 |
| *SMAD7* | SMAD family member 7; MIM * 602932 |
| SMC | Smooth muscle cell |
| SNP | Single nucleotide polymorphism |
| SNV | Single nucleotide variant |
| SV | Structural variant |
| T | Thymine |
| TAA | Thoracic aortic aneurysm |
| TAAD | Familial thoracic aortic aneurysms and dissections; MIM #132900 |
| TβRI | TGFβ type I receptor |
| TβRII | TGFβ type II receptor |
| TGFβ | Transforming growth factor beta |
| *TGFB1* | Transforming growth factor, beta 1; MIM *190180 |
| *TGFB2* | Transforming growth factor, beta 2; MIM *190220 |
| *TGFB3* | Transforming growth factor, beta 3; MIM *190230 |
| *TGFBR1* | Transforming growth factor, beta receptor 1; MIM *190181 |
| *TGFBR2* | Transforming growth factor, beta receptor II (70/80kDa); MIM *190182 |
| *TGFBR3* | Transforming growth factor, beta receptor III; MIM *600742 |
| *THBS2* | Thrombospondin 2; MIM *188061 |
| *THSD4* | Thrombospondin, type I, domain containing 4; MIM *614476 |
| *TNXB* | Tenascin XB; MIM *600985 |
| *TSC2* | Tuberous sclerosis 2; MIM *191092 |
| UPD | Uniparental disomy |
| *VCAN* | Versican; MIM *118661 |
| VCF | Variant Call Format |
| VSMC | Vascular smooth muscle cell |
| *VWA3A* | Von Willebrand factor A domain containing 3A |
| WES | Whole-exome sequencing |
| WGS | Whole-genome sequencing |

# 1 General Introduction

## 1.1 Aortic Diseases

### 1.1.1 Cardiovascular System

The cardiovascular system is responsible for the transport of nutrients and oxygen throughout the body. The motor of the cardiovascular system is the heart, which is divided into two halves, each consisting of an atrium and a chamber (Figure 1). The right half receives oxygen poor blood from the body and pumps it into the lungs for oxygenation. From there it enters the left half of the heart via the pulmonary veins and is pumped back to the body through the aorta to supply the body with oxygen. Thus, the aorta is the blood vessel which experiences the highest blood pressures and is consequently most sensitive to changes in the connective tissue composition within its wall [reviewed in Quaglino and Ronchetti 2002].

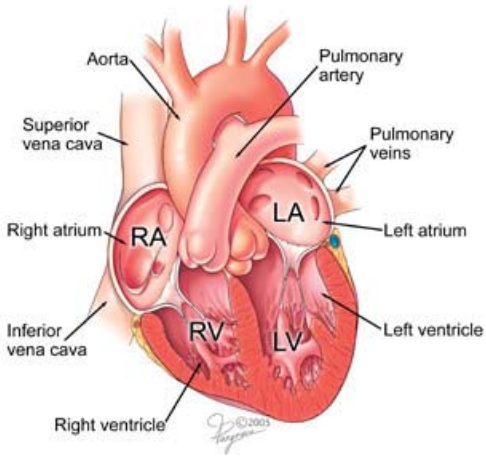



**Figure 1**. Structural composition of the human heart. LA, left atrium; LV, left ventricle; RA, right atrium; RV, right ventricle (http://www.stopafib.org/images/diagram-heart-pump-large.jpg).

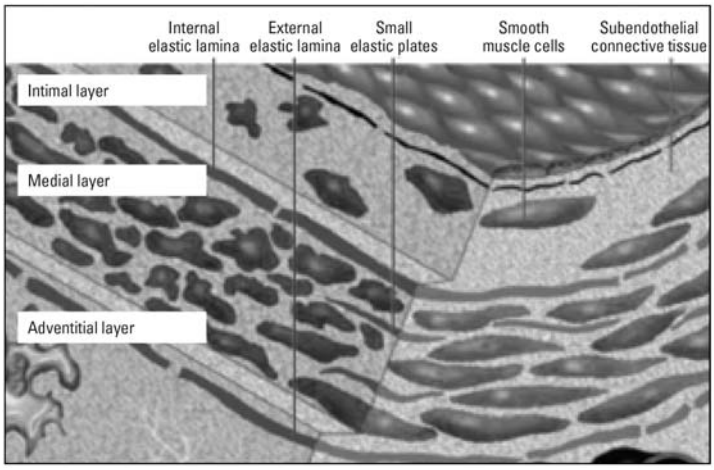



**Figure 2**. The three layers of vessel walls and their composition [Brunner and Ignaszewski 2011].

The walls of blood vessels consist of three different layers (Figure 2). While the *tunica intima* (intimal layer) consists of a single layer of endothelial cells, which forms the endothelium, the *tunica media* (medial layer) is characterized by alternating layers of vascular smooth muscle cells (VSMCs) and concentric elastic lamellae. The number of these layers is highly variable according to the function of the vessel. Likewise, the ascending aorta which is exposed to the highest pressure contains most of these layers and the number is decreasing according to the reduction of maximal pressure with increasing distance from the heart. Also some collagen bundles and proteoglycans can be seen among the VSMCs. In contrast, the *tunica adventitia* (adventitial layer) is composed of dense connective tissue mainly comprised of collagen bundles and only a few spread elastic aggregates. Fibroblasts are the most abundant cell type in this part of the vessel wall, but macrophages can also

appear. The constitution of the layers is different depending on vessel size and type (Table 1) [reviewed in Quaglino and Ronchetti 2002].

VSMCs in the *tunica media* of the aorta, which belongs to the elastic arteries (Table 1), are produced from at least seven unique and non-overlapping origins in vertebrate embryos depending on their location in the aorta (Figure 3). VSMCs from different embryonic origins respond in lineage-specific ways to important soluble factors that control development, growth, and remodelling of the vessel wall and may thus be variably sensitive to changes in signalling, like in the case of inherited aortic diseases (AD) [reviewed in Majesky 2007].

**Table 1**. Vessel parameters in humans [Quaglino and Ronchetti 2002].

| | **Type** | **Diameter** | ***Tunica Intima*** | ***Tunica Media*** | ***Tunica Adventitia*** |
|---|---|---|---|---|---|
| **Arteries** | | | | | |
| | Elastic | >1 cm | 50-100 µm thick (increases with age), endothelium, basement membrane, proteoglycans, a few collagen fibrils, internal elastic lamina | 1-3 mm thick, layers of smooth muscle cells, layers of elastic lamellae, small collagen bundles, proteoglycans | 300-400 µm thick, fibroblasts, scarce smooth muscle cells and elastic fibres, thick collagen bundles, vasa vasorum |
| | Muscular | 0.1-10 mm | 50-100 µm thick (increases with age), endothelium, basement membrane, proteoglycans, a few collagen fibrils and smooth muscle cells, prominent internal elastic lamina | 0.1-1 mm thick, layers of smooth muscle cells, scarce elastic lamellae, small collagen bundles, proteoglycans | 200-400 µm thick, fibroblasts, scarce smooth muscle cells and elastic fibres, thick collagen bundles, vasa vasorum |
| | Arterioles | <100 µm | Thin endothelium, basement membrane, little collagen, thin elastic lamina | One to two layers of smooth muscle cells, scarce elastin, collagen, and proteoglycans | Thin, ill-defined sheets of connective tissue |
| | Capillaries | <10 µm | Endothelium, basement membrane | None | None |
| **Veins** | | | | | |
| | Postcapillaries | 10-30 µm | 5-20 µm thick (increases with age), endothelium, basement membrane, pericytes | None | None |
| | Small | 0.1-1 mm | 20-50 µm thick, endothelium, basement membrane, pericytes, rare smooth muscle cells, scarce collagen fibrils and proteoglycans | 50-100 µm thick, smooth muscle cells (one to three layers), scarce collagen bundles and proteoglycans | 100-200 µm thick, fibroblasts, collagen bundles, some elastic fibres |
| | Medium | 1-10 mm | 20-50 µm thick, endothelium, basement membrane, rare smooth muscle cells, scarce collagen fibrils and proteoglycans | 100-500 µm thick, smooth muscle cells (up to eight layers), scarce collagen bundles, elastic fibres and proteoglycans | 200-500 µm thick, fibroblasts, scarce smooth muscle cells, collagen bundles, some elastic fibres |
| | Large | >1 cm | 20-50 µm thick, endothelium, basement membrane, rare smooth muscle cells, scarce collagen fibrils and proteoglycans | >0.5 mm thick, few fibroblasts, smooth muscle cells (up to 15 layers), collagen bundles, elastic fibres, proteoglycans | >0.5 mm thick, fibroblasts, smooth muscle cells, collagen bundles, some elastic fibres |

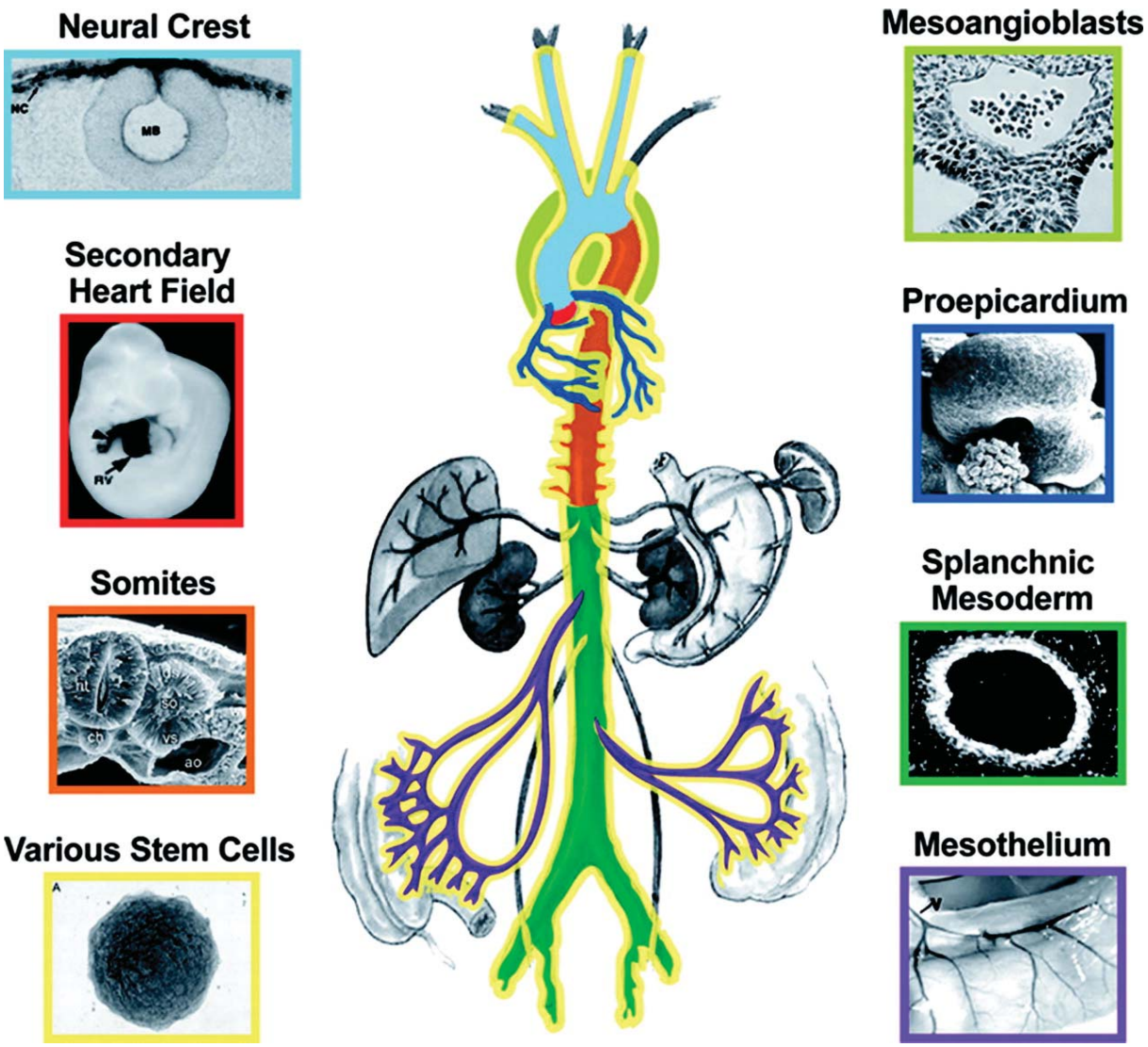



**Figure 3.** Developmental fate map for vascular smooth muscle cells (VSMCs). Different colours represent different embryonic origins for VSMCs as indicated in the boxed images to the left and right sides of the figure. Yellow outline indicates both local and systemic contributions by various sources of vascular stem cells. The fate map reveals a highly mosaic distribution of VSMC subtypes in the aorta and its major branch arteries [Majesky 2007].

### 1.1.2 Inherited Syndromic and Non-Syndromic Aortic Diseases

Aortic aneurysms and dissections are associated with high morbidity and mortality due to an increased risk for aortic rupture. Aneurysms are dilatations of blood vessels, which result in thinner and less stable vessel walls. Two main types of aortic aneurysms are distinguished depending on their location in the aorta. Aneurysms above the diaphragm are classified as thoracic aortic aneurysms (TAA), whereas such below the diaphragm are known as abdominal aortic aneurysms (AAA) (Figure 4). Aortic aneurysms tend to be asymptomatic and are often only diagnosed in the event of aortic dissection or rupture. Aortic dissection is a tear in the intimal layer of the aortic wall and can occur with or without prior aortic aneurysm (Figure 2, Figure 5). Blood from the aortic lumen enters the vessel wall through this tear and dissects along the plane of the wall establishing a false lumen. In the classification of

Stanford aortic dissections are divided according to whether the ascending aorta is involved (type A) or not involved (type B) [reviewed in Milewicz *et al*. 2008 and Nienaber and Clough 2015]. Aortic dissection not only leads to a substantial risk for aortic rupture, but can also cause obstruction of flow to aortic branch vessel by an intimal flap leading to malperfusion of the vascular territory of affected vessels [Swee and Dake 2008]. Risk factors for aortic diseases include age, male gender, smoking, high blood pressure, and especially in the case of younger individuals with thoracic aortic aneurysms and dissections also bicuspid aortic valves and genetic factors [Coady *et al*. 1999, Landenhed *et al*. 2015].

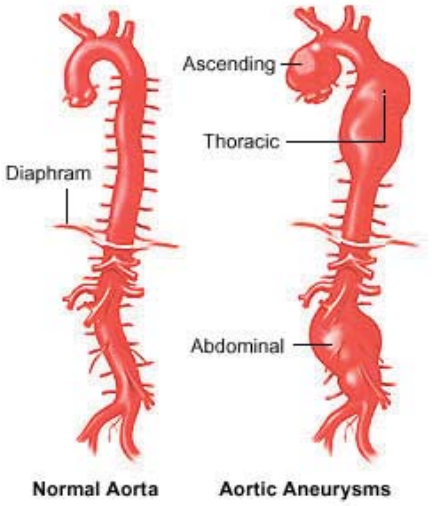



**Figure 4**. Classification of aortic aneurysms (http://www.northernsydneyvascular.com.au/images/AorticAneurysm1.jpg).

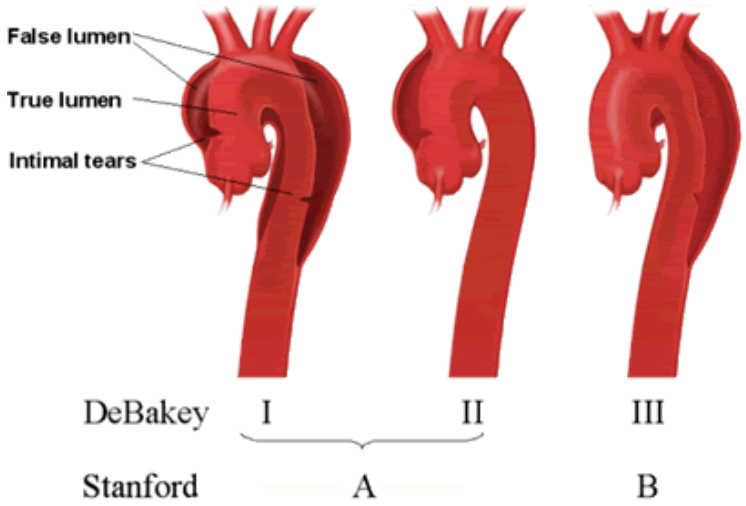
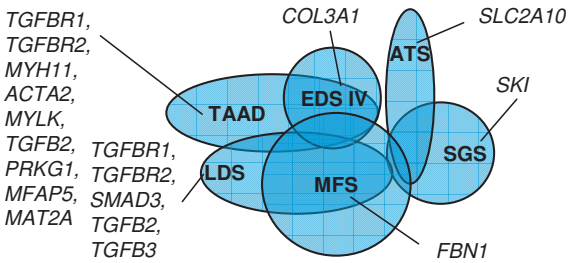



**Figure 5**. Classification of aortic dissections (http://img.photobucket.com/albums/v671/passmore/AD2.gif).




**ATS**: Arterial tortuosity syndrome, ar
**EDS IV**: Ehlers-Danlos syndrome vascular type, ad
**LDS**: Loeys-Dietz syndrome, ad
**MFS**: Marfan syndrome, ad
**SGS**: Shprintzen-Goldberg syndrome, ad
**TAAD**: Familial thoracic aortic aneurysms and dissections, ad

**Figure 6**. Inherited disorders associated with aortic disease and overlapping phenotypes. ad, autosomal dominant; ar, autosomal recessive [kindly provided by G. Matyas].

Aortic aneurysms and dissections can occur in a broad range of inherited (connective tissue) disorders with overlapping clinical signs (Figure 6, Table 2). However, only a limited part of genes mutated in AD is known so far. These comprise the non-syndromic familial thoracic aortic aneurysms and dissections (TAAD) caused by mutations in *TGFBR1* and *TGFBR2* [Pannu *et al*. 2005, Matyas *et al*. 2006], *MYH11* [Zhu *et al*. 2006], *ACTA2* [Guo *et al*. 2007], *MYLK* [Wang *et al*. 2010], *TGFB2* [Boileau *et al*. 2012], *PRKG1* [Guo *et al*. 2013], *MFAP5* [Barbier *et al*. 2014], and *MAT2A* [Guo *et al*. 2015]. In addition, also syndromic disorders such as Marfan syndrome (MFS) associated with mutations in *FBN1* [Dietz *et al*. 1991], Ehlers-Danlos syndrome vascular type (EDS IV) associated with mutations in *COL3A1* [e.g. Pope *et al*. 1975, Superti-Furga *et al*. 1988, Tromp *et al*. 1989, Steinmann *et*

*al*. 2002], Loeys-Dietz syndrome (LDS) caused by mutations in *TGFBR1* and *TGFBR2* [Loeys *et al*. 2005, 2006], *SMAD3* [van de Laar *et al*. 2011, 2012], as well as *TGFB2* [Boileau *et al*. 2012, Lindsay *et al*. 2012], arterial tortuosity syndrome (ATS) due to mutations in *SLC2A10* [Coucke *et al*. 2006], and Shprintzen-Goldberg syndrome (SGS) associated with mutations in *SKI* [Doyle *et al*. 2012a] have been described. Recently heterozygous mutations in *TGFB3* have been reported in individuals with clinical features overlapping MFS and LDS [Rienhoff *et al*. 2013, Matyas *et al*. 2014, Bertoli-Avella *et al*. 2015].

**Table 2**. Disorders associated with AD [Attenhofer Jost *et al*. 2014].

| Disorder/syndrome | Inheritance | Prevalence (incidence) | Aortic aneurysm | Early aortic dissection | Arterial tortuosity | Other cardiovascular features | Gene (karyotype) | Pathway |
|---|---|---|---|---|---|---|---|---|
| Marfan syndrome | ad | ~1:5,000 | ++ | + | - | IA, MVP | *FBN1* | TGF-ß |
| TGFBR1/TGFBR2-related Loeys-Dietz syndrome (LDS) and thoracic aortic aneurysm/dissection (TAAD) | ad | unknown | ++ | +++ | ++ | BAV, IA, MVP | *TGFBR1, TGFBR2* | TGF-ß |
| SMAD3-related LDS, TAAD, and aneurysms-osteoarthritis syndrome | ad | unknown | ++ | ++ / +++ | ++ | BAV, IA, MVP | *SMAD3* | TGF-ß |
| TGFB2-related LDS and TAAD | ad | unknown | ++ / +++ | + | + | BAV, MVP | *TGFB2* | TGF-ß |
| ACTA2-, MYH11-, and MYLK-related TAAD | ad | unknown | +++ | ++ | - | BAV, CAD (*ACTA2*), PDA | *ACTA2, MYH11, MYLK* | IGF-1, ANG-II |
| Ehlers-Danlos syndrome, vascular type (EDS IV) | ad | ~1:50,000 | + | ++ (rupture) | + | IA, MVP | *COL3A1* | collagen metabolism |
| Ehlers-Danlos syndrome, kyphoscoliotic form (EDS VIA) | ar | (~1:100,000) | + | ++ (rupture) | - | MVP | *PLOD1* | collagen metabolism |
| PTPN11-related Noonan and LEOPARD syndromes | ad | ~1:2,000 | + | + | - | pulmonary valve stenosis | *PTPN11* | RAS-MEK-ERK |
| JAG1-related Alagille syndrome | ad | (1:70,000) | + | + | - | pulmonary valve stenosis, COA, IA, TOF | *JAG1* | NOTCH1-JAGGED1 |
| Aortic valve disease | ad | unknown | + | + | - | BAV with valve calcification/dysfunction | *NOTCH1* | NOTCH1-JAGGED1 |
| Congenital Contractural Arachnodactyly | ad | unknown | + | - | - | atrial/ventricular septal defects, MVP | *FBN2* | TGF-ß |
| SKI-related Shprintzen-Goldberg Syndrome | ad | unknown | ++ | - | + | MVP, splenic artery aneurysm | *SKI* | TGF-ß |
| ELN-related cutis laxa | ad | (<1:4,000,000) | + | + | - | - | *ELN* | unknown |
| EFEMP2-related cutis laxa | ar | (<1:4,000,000) | ++ | + | ++ | arterial stenoses | *EFEMP2* | TGF-ß |
| Arterial tortuosity syndrome | ar | unknown | + | + | +++ | arterial stenoses | *SLC2A10* | TGF-ß |
| FLNA-related periventricular heterotopia | Xld | unknown | + | + | - | BAV, PDA | *FLNA* | unknown |
| Fabry disease, cardiac variant | Xl | (~1:3,000) | + | + | + | HCM | *GLA* | unknown |
| X-linked Alport syndrome | Xl | (<1:50,000) | + | + | - | - | *COL4A5* | collagen metabolism |
| Turner syndrome | sporadic | (1:2,000) | + | ++ | - | BAV, COA, IA, LVOTO | (45,X) | unknown |

Extremely rare aortic aneurysm (AA) is caused by mutations in the genes *COL1A1*, *COL1A2* or *MED12*, whereas mutations in the genes *PLOD3*, *ENG*, *ACVRL1* or *NF1* cause medium-sized AAs. BAV, bicuspid aortic valve; CAD, coronary artery disease; COA, coarctation of the aorta; HCM, hypertrophic cardiomyopathy; IA, intracranial aneurysms; LVOTO, left ventricular outflow tract obstruction; MVP, mitral valve prolapse; PDA, patent ductus arteriosus; TOF, tetralogy of Fallot; -, absent or not observed/reported; +, sporadic; ++, common; +++, typical; ad, autosomal dominant; ar, autosomal recessive; Xl, X-linked; Xld, X-linked dominant.

## 1.1.3 Signalling Pathways Involved in Aortic Diseases

### 1.1.3.1 TGFβ Signalling

The transforming growth factor beta (TGFβ) superfamily consists of cytokines that regulate a wide range of functions like cell proliferation, differentiation, migration, adhesion, apoptosis, and extracellular matrix (ECM) production. Consequently, it plays a crucial role in the pathogenesis of different diseases such as cancer, autoimmune diseases, fibrosis, and cardiovascular diseases. There are at least three isoforms of TGFβ (TGFβ1-3), which are highly conserved proteins synthesized as large protein precursors consisting of an N-terminal signal peptide, a pro region (latency-associated peptide, LAP), and a C-terminal portion. This C-terminal portion is cleaved from the LAP within the cell to become the mature TGFβ molecule. The LAP and the mature TGFβ molecule remain non-covalently bound building the small latent complex (SLC) and keeping TGFβ in an inactive state. Following binding to one of four latent TGFβ-binding proteins (LTBP1-4) forming the large latent complex (LLC), it is secreted from the cell and sequestered in fibrillin-rich microfibrils of the ECM through interaction of LTBP with the N-terminus of fibrillin-1. TGFβ gets activated when the free ligand is released from LLC by proteolytic cleavage by thrombospondin, plasmin, reactive oxygen species, acidic microenvironment, matrix metalloproteinases (MMP2 and MMP9) or β6 integrin. Upon activation TGFβ binds to a homodimer of TGFβ type II receptors (TβRIIs), constitutively active serine/threonine kinase transmembrane receptors, which then recruits a TGFβ type I receptor (TβRI) homodimer and activates it by phosphorylation of specific serine and threonine residues in its intracellular juxtamembrane region. Activated TβRI, which also belongs to the family of serine/threonine kinase transmembrane receptors, recruits and phosphorylates receptor-regulated SMAD (R-SMAD), namely SMAD2 and/or SMAD3. These activated R-SMADs bind to the common SMAD (co-SMAD) SMAD4 to form a heterodimeric complex, which enters the cell nucleus where it regulates together with other transcription factors the expression of target genes. In addition to this highly conserved so-called canonical (i.e. SMAD-dependent) pathway, TGFβ receptor activation leads also to stimulation of several non-SMAD (non-canonical) signalling cascades, including Rho-associated protein kinase (ROCK) and mitogen-activated protein kinase (MAPK) cascades, the latter of which includes extracellular signal-regulated kinase (ERK), c-Jun N-Terminal kinase (JNK), and p38 (Figure 7) [reviewed in Doyle *et al*. 2012b and Pardali and ten Dijke 2012].

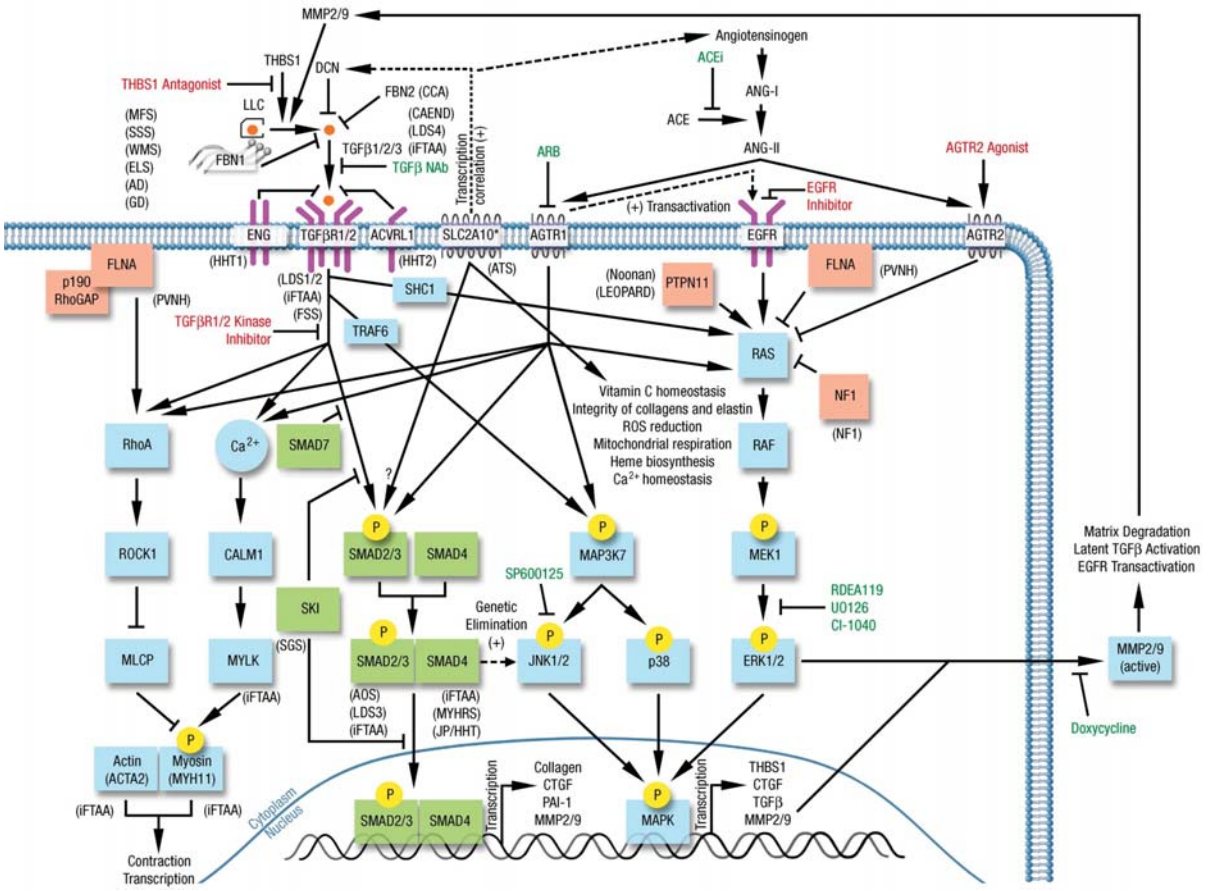



**Figure 7**. Signalling pathways involved in AD. Green indicates canonical TGFβ signalling, blue non-canonical TGFβ signalling, and red other proteins implicated in TGFβ signalling [Attenhofer Jost *et al*. 2014].

It was long thought that MFS caused by mutations in the gene *FBN1* is a disorder of weak connective tissue like EDS IV, which is caused by mutations in the *COL3A1* gene encoding the pro-alpha 1 chains of type III collagen, a fibrillar collagen of extensible connective tissues. However, some features of MFS like bone overgrowth, may not solely be explained by the deficiency of an ECM component like fibrillin-1, the protein encoded by *FBN1*. In 2003, Neptune *et al*. described that *FBN1* mutations may also lead to enhanced TGFβ signalling [Neptune *et al*. 2003], which may explain these features. It was proposed that disturbed microfibril structure due to *FBN1* mutations may lead to lower amount of bound LLC and consequently less TGFβ can be kept in its inactive form resulting in enhanced TGFβ signalling. Later it has been shown that enhanced TGFβ signalling also contributes to other features of MFS such as the increased risk for aortic aneurysms and dissections [reviewed in Doyle *et al*. 2012b].

A similar mechanism is proposed for mutations in *MFAP5*, which have very recently been associated with TAAD and which were shown to lead to enhanced TGFβ signalling in aortas of affected patients [Barbier *et al*. 2014]. This gene encodes microfibril-associated glycoprotein-2 (MAGP-2) known to interact with fibrillin-1 of isolated or elastin-associated microfibrils [Gibson *et al*. 1998]. *MAT2A*, loss-of-function mutations in which have very recently been reported to predispose to aortic aneurysms and bicuspid aortic valve, encodes

the methionine adenosyltransferase MAT2α. This enzyme is involved in the synthesis of the amino acid cysteine and loss-of-function mutations lead thus to reduced intracellular cysteine amounts, which could affect the synthesis of the cysteine-rich protein fibrillin-1. Alternatively or in addition, also increased oxidative stress enhancing signalling through the angiotensin II pathway, which activates similar signalling cascades as TGFβ (Figure 7), could contribute to the predisposition to AD [Guo *et al*. 2015]. Another player that binds the inactive form of TGFβ is decorin. Decorin is expressed in a glucose-dependent manner. Likewise, mutations in *SLC2A10* encoding the glucose transporter GLUT10, which is localised in the mitochondrial walls of VSMCs, are associated with autosomal recessively inherited ATS. This disorder is also characterized by enhanced TGFβ signalling and has features overlapping with MFS and LDS including the increased risk for aneurysms and dissections of large arteries [Coucke *et al*. 2006].

Further confirmation for the role of aberrant TGFβ in AD was the association of mutations in genes encoding proteins directly involved in TGFβ signalling with syndromic forms of AD including marfanoid habitus. After the association of mutations in the receptors (*TGFBR1* and *TGFBR2*) with LDS [Loeys *et al*. 2005, 2006], also mutations in *SMAD3*, which is one of the R-SMADs involved in canonical TGFβ signalling, were associated with a related phenotype [van de Laar *et al*. 2011, 2012]. Recently, also mutations in the ligands (*TGFB2* and *TGFB3*) have been associated with LDS [Boileau *et al*. 2012, Lindsay *et al*. 2012, Rienhoff *et al*. 2013, Matyas *et al*. 2014, Bertoli-Avella *et al*. 2015]. Also mutations directly in an inhibitor of TGFβ signalling are known to cause a related phenotype, like in the case of SGS, which is associated with mutations in the gene *SKI* encoding a negative regulator of SMAD-dependent TGFβ signalling [Doyle *et al*. 2012a]. While first studies indicated that AD is mainly mediated by canonical TGFβ signalling, recent results with mouse models also indicated an important role of non-canonical TGFβ signalling in the pathogenesis of AD (Figure 7) [Habashi *et al*. 2011, Holm *et al*. 2011].

### 1.1.3.2 Contractile Apparatus

Interestingly, many genes exclusively associated with TAAD without systemic features encode proteins involved in the contractile apparatus of VSMCs and are not directly linked to TGFβ signalling. Likewise, *ACTA2* and *MYH11*, mutations in both of which are associated with TAAD, encode the most abundant actin and myosin heavy chains in VSMCs, respectively [Zhu *et al*. 2006, Guo *et al*. 2007]. *MYLK*, a further gene associated with TAAD, encodes the myosin light chain kinase (MLCK), which initiates the physiological contractions of smooth muscle cells (SMCs) in hollow organs such as the aorta [Wang *et al*. 2010]. Mutations in these three genes are associated with a defective contractile apparatus. So far, only one single recurrent gain-of-function mutation in *PRKG1* leading to a constitutively

active enzyme has been associated with TAAD [Guo *et al*. 2013]. As this gene encodes type I cGMP-dependent protein kinase (PKG-1), which activates SMC relaxation, it is still in line with the proposed disease mechanism of decreased SMC contraction leading to TAAD. Another proposed mechanism of how mutations in these genes of the contractile apparatus predispose to aortic aneurysm is reduced mechanosensing. Accordingly, cells will mistake high stress for low stress and thus activate degenerative pathways [Humphrey *et al*. 2014].

### 1.1.4 Current Therapeutic Possibilities

Inherited AD cannot be cured but managed. Standard management/treatment includes blood pressure control and adaptation of life style, which includes stress avoidance and no contact sport or other activities raising the blood pressure. Another therapeutic option is elective aortic surgery in order to avoid emergency situations, which are associated with a worse outcome. The decision for elective surgery depends on different factors like aortic diameter, progression rate, and underlying genetic defect. Likewise, for LDS patients, elective surgery is recommended at lower diameter than in MFS (Table 3) [reviewed in Attenhofer Jost *et al*. 2014]. The standard medical therapy to reduce the rate of aortic enlargement in MFS patients are β-blockers despite limited evidence of beneficial effects [Shores *et al*. 1994, Attenhofer Jost *et al*. 2014]. Comparatively, a recent clinical trial with EDS IV patients has shown a beneficial effect of the β-blocker celiprolol to prevent arterial ruptures and dissections [Ong *et al*. 2010].

**Table 3**. Current guidelines for operative aortic root replacement [Attenhofer Jost *et al*. 2014].

| Disorder/syndrome | Aortic root replacement if size is larger than* |
|---|---|
| Marfan syndrome | 5 cm |
| TGFBR1/TGFBR2-related Loeys-Dietz syndrome (LDS) and thoracic aortic aneurysm/dissection (TAAD) | 4-4.2 cm (internal diameter) or 4.4-4.6 cm CT and/or MRI |
| SMAD3-related LDS, TAAD, and aneurysms-osteoarthritis syndrome; TGFB2-related LDS and TAAD | No data, may similar to LDS |
| ACTA2-, MYH11-, and MYLK-related TAAD; Ehlers-Danlos syndrome, vascular type (EDS IV); Ehlers-Danlos syndrome, kyphoscoliotic form (EDS VIA); PTPN11-related Noonan and LEOPARD syndromes; JAG1-related Alagille syndrome | No data |
| Turner syndrome | 4.5-5 cm (aorta >2.5 cm/m$^2$) |

*Comment: Earlier prior to pregnancy, positive family history for aortic dissections, rapid growth of the aorta (>5 mm/year), associated aortic valve disease, maximum aortic cross-sectional/area/body height >10 cm$^2$/m.

**Table 4**. Overview of clinical trials on pharmacological treatment of dilatation of the aorta in Marfan syndrome (ARB, β-blocker or both) [Attenhofer Jost *et al*. 2014].

| Trial identifier (reference) | Year trial started | Goal (study design) | Number of patients, time follow-up | Age group included | β-blocker (sample size), dosing | ARB (sample size), dosing | Outcome parameter |
|---|---|---|---|---|---|---|---|
| NCT00429364 (Lacro *et al.* [54]) | 2007 | Comparing losartan vs. atenolol (randomized, single blind) | 604 MFS patients, 3 years follow-up | 6 months to 25 years and aortic root z-score >3.0 | Atenolol (302), 0.5-4 mg/kg/day max. 250 mg/day | Losartan (302), 0.3-1.4 mg/kg/day max. 100 mg/day | Rate of change in aortic root BSA adjusted z-score |
| NCT00651235 (Chiu *et al.* [38] | 2007 | Efficacy of losartan added to β-blocker (randomized, open label) | 28 MFS patients, 35 months of follow-up | ≥1 year and aortic root z-score ≥2.0 | β-blockers (13), atenolol or propranolol max. 150 mg/day for adults and 2 mg/kg/day for children | Losartan and β-blocker (15), 100 mg/day for adults and 50 mg/day for children | Change in aortic root diameter |
| NCT00782327 (Möberg *et al.* [36]) | 2009 | Additive effect of losartan and β-blocker (randomized, double blind) | 174 MFS patients, 3 years follow-up | ≥10 years and aortic root z-score ≥2 | β-blocker and placebo (87), no data on dosing | Losartan and β-blocker (87), 25-50 mg/day below 50 kg or 50-100 mg/day over 50 kg | Decrease in aortic root growth rate |
| NCT00763893 (Detaint *et al.* [34]) | 2008 | Efficacy of losartan vs. placebo (randomized, double blind) | 300 MFS patients, 3 years follow-up | ≥10 years | No β-blocker | Losartan (150) and placebo (150), 50 mg/day below 50 kg or 100 mg/day over 50 kg | Change in aortic root diameter |
| NTR1423 (Radonic *et al.* [35]) | 2008 | Efficacy of losartan vs. not-treated controls (randomized, open label) | 330 MFS patients, 3 years follow-up | ≥18 years | No β-blockers (165) but patients continue taking their standard β-blocker treatment | Losartan (165), 50 mg/day (0-14 days) or 100 mg/day (>14 days) | Change in aortic root diameter and skin gene expression |
| NCT01145612 (Forteza *et al.* [55]) | 2008 | Efficacy of losartan vs. atenolol (randomized, double blind) | 150 MFS patients, 3 years follow-up | 5-60 years and aortic diameter <45 mm | Atenolol (75), 12.5 mg/day (0-14 days) and 25 mg/day (>14 days) below 50 kg or 25 mg/day (0-14 days) and 50 mg/day (>14 days) over 50 kg | Losartan (75), 12.5 mg/day (0-14 days) and 25 mg/day (>14 days) below 50 kg or 25 mg/day (0-14 days) and 50 mg/day (>14 days) over 50 kg | Progression of dilation of the aortic valve annulus, sinuses of Valsalva, sinotubular junction, ascending aorta, aortic arch, thoracic and abdominal aorta |
| NCT00683124 (Gambarin *et al.* [33]) | 2008 | Effects of losartan vs. nebivolol vs. the combination of both (randomized, open label) | 291 MFS patients with FBN1 mutation, 4 years follow-up | 12 months to 55 years and aortic root z-score ≥2.5 but <50 mm | Nebivolol (97 + 97 in combination?), max. 10 mg/day for adults and max. 0.16 mg/kg/day for children <16 years | Losartan (97 + 97 in combination?), max. 100 mg/day for adults and max. 1.6 mg/kg/day for children <16 years | BSA and age-adjusted aortic root diameter (sinuses of Valsalva), drug responsiveness (losartan: CYP2C9 gene, nebivolol: CYP2D6 gene) |
| NCT00723801 (ClinicalTrials.gov) | 2007 | Effects of losartan vs. atenolol on aortic stiffness (randomized, double blind) | 50 MFS patients, 6 months follow-up | ≥25 years | Atenolol (25?), 50 mg/day | Losartan (25?), 100 mg/day | Aortic biophysical properties and diastolic function |

NCT, ClinicalTrials.gov Identifier; NTR, Netherlands trial register; MFS, Marfan syndrome

The knowledge on the involvement of high TGFβ signalling in AD also opened new therapeutic possibilities. In a mouse model of MFS, it was shown that the reduction of TGFβ signalling using antibodies against TGFβ slowed down the growth rate of aortic aneurysms. Likewise, also the administration of angiotensin II type 1 receptor blockers (ARBs) like losartan has been shown to be efficient in MFS mice [Habashi *et al*. 2006]. A small study with 18 young adults showed the efficacy of losartan to reduce the aortic growth rate also in humans [Brooke *et al*. 2008]. Larger clinical trials with different age groups and study protocols have consequently been initiated to further assess the efficacy of ARBs also compared or in addition to the current standard treatment with β-blockers (Table 4) [reviewed in Attenhofer Jost *et al*. 2014]. First results of these trials confirm that losartan treatment

significantly reduced aortic growth rate and show that a combination of β-blockers and losartan is more efficient to prevent a fatal cardiac outcome than β-blockers alone [Chiu *et al*. 2013, Groenink *et al*. 2013, Pees *et al*. 2013, Mueller *et al*. 2014]. However, recently published results of the clinical trial of the Pediatric Heart Network detected no significant differences in the reduction of root growth in groups either treated with the ARB losartan or the β-blocker atenolol [Lacro *et al*. 2014]. Studies with mouse models of LDS, which is like MFS associated with enhanced TGFβ signalling despite truncating mutations in active players of this pathway, have also shown that losartan in contrast to the β-blocker propranolol is effective to reduce TGFβ signalling and ameliorate pathologic aortic growth, suggesting ARBs also as a treatment option for LDS [Gallo *et al*. 2014].

Angiotensin-converting enzyme inhibitors (ACEIs) inhibit the production of angiotensin II and thus signalling through both angiotensin II receptors (ATR1 and ATR2). While signalling through ATR1 activates ERK1/2, which is involved in non-canonical TGFβ signalling, it is inhibited by signalling through ATR2. Likewise, in mouse models of MFS direct inhibition of ERK1/2 as well as treatment with the ARB losartan, which selectively inhibits ATR1, were shown to be more efficient to reduce AD than treatment with ACEI [Holm *et al*. 2011, Habashi *et al*. 2011].

A different therapeutic approach is the tetracyclic antibiotics doxycycline, which has been shown to ameliorate the aortic phenotype in mouse models of MFS and EDS IV [Xiong *et al*. 2008, Briest *et al*. 2011, Tae *et al*. 2012]. This drug has in subantibiotic concentration an inhibitory effect on MMPs, which degrade ECM components such as fibrillin and collagen. Another approach targeting matrix degradation through MMPs is the application of pravastatin, which was reported to reduce aortic root dilatation in MFS mouse models [McLoughlin *et al*. 2011]. Pravastatin is marketed as Pravachol or Selektine, belongs to the drug class of statins, and is normally used in combination with diet, exercise, and weight loss for lowering cholesterol and preventing cardiovascular disease like atherosclerosis. In this case, however, it acted as an inhibitor of isoprenoids, which are essential in the transportation and secretion of MMPs, leading to decreased matrix degeneration by MMPs.

## 1.2 Novel High-Throughput Technologies in Molecular Genetics

### 1.2.1 Microarrays

Traditional cytogenetics using caryotyping can only detect chromosome aberrations larger than 5-10 Mb. The gap for detection of deletions and duplications in the size range of 1 kb-10 Mb is closed by microarrays, of which two main types are available. One approach are SNP arrays, which contain probes for common single nucleotide polymorphisms (SNPs) to which labelled DNA of one sample is hybridized (i.e. one DNA sample per array). The most common applications of these arrays are genome-wide association studies (GWAS), which associate common sequence variants with phenotypic traits. Alternatively, the information of SNP arrays can be used to detect regions with abnormal allele distribution. Loss of heterozygositiy (LOH) in combination with decreased signal intensities indicates deletions, whereas copy neutral LOH indicates uniparental disomy (UPD) or consanguinity. Duplications are detected as increased signal intensity and relative allele rations different from 0.5 at heterozygous positions. In contrast to SNP arrays in array comparative genomic hybridization (aCGH) differently labelled control and patient DNA is loaded on the same array and the relative signal strength determines the copy number (Figure 8). Depending on the density of probes the resolution of aCGH can be up to ~400 bp. Both of these array types allow only the detection of copy number variants (CNVs), i.e. large deletions and duplications, but not of copy neutral structural variants (SVs) such as insertions and balanced translocations [reviewed in Le Scouarnec and Gribble 2012 and Riegel 2014]. For exact breakpoint localisation on base pair level additional methods like long-range polymerase chain reaction (LR-PCR) in combination with Sanger sequencing are needed, which are laborious and time consuming [Matyas *et al*. 2007, Meienberg *et al*. 2010 (Appendix 1)].

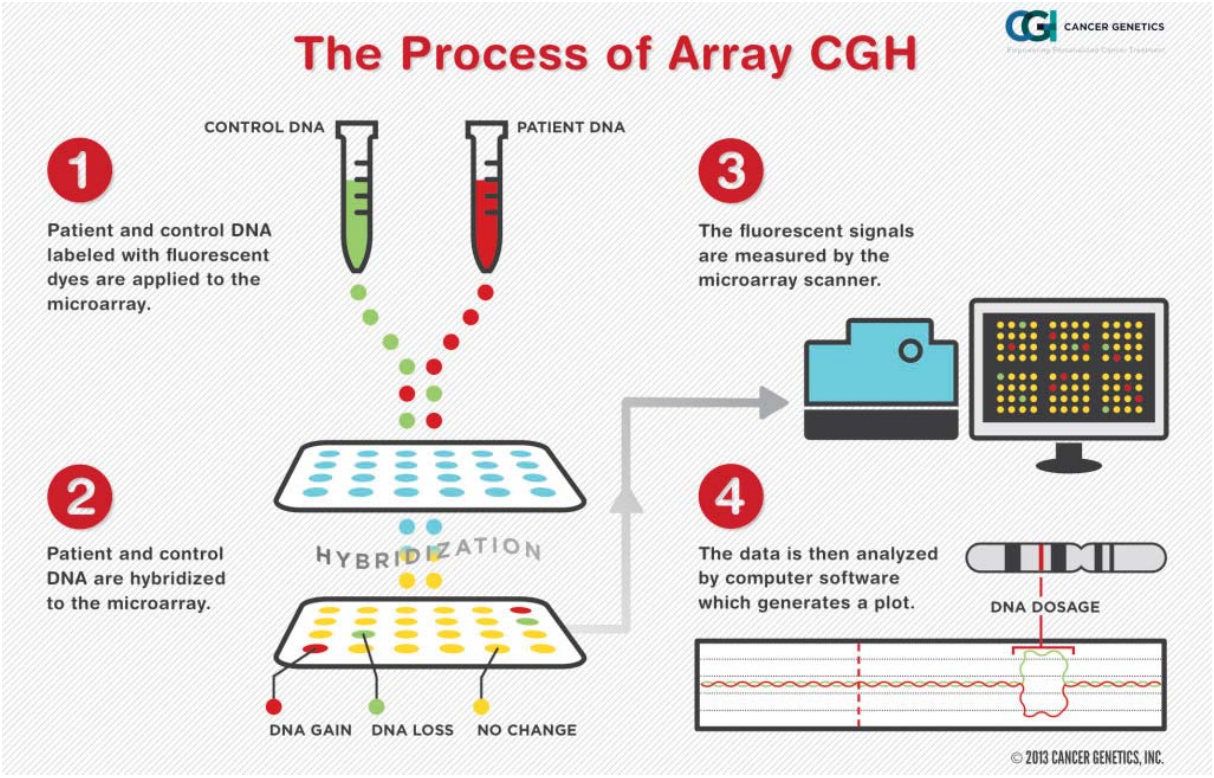



**Figure 8**. The general procedure of aCGH (http://cgimatba.com/array-cgh-technology).

### 1.2.2 Next-Generation Sequencing

The current gold standard in sequencing is the strand termination method by Sanger [Sanger *et al*. 1977]. This method allows to sequence up to 1,000 bp of a defined region in one reaction. In contrast, massively parallel sequencing, also referred to as next-generation sequencing (NGS), which is commercially available since 2005, allows sequencing of many different fragments within one reaction. Using NGS, it is much less laborious to sequence high numbers of fragments allowing a much higher throughput and lower costs per sequenced base compared to traditional Sanger sequencing. Accordingly, the time and costs to sequence a complete human genome dropped from 13 years and estimated costs of $2.7 billion for the first human genome completed in 2003 through the Human Genome Project using traditional Sanger sequencing to a few days and less than $10,000 using NGS [reviewed in Voelkerding *et al*. 2009 and Chrystoja and Diamandis 2014].

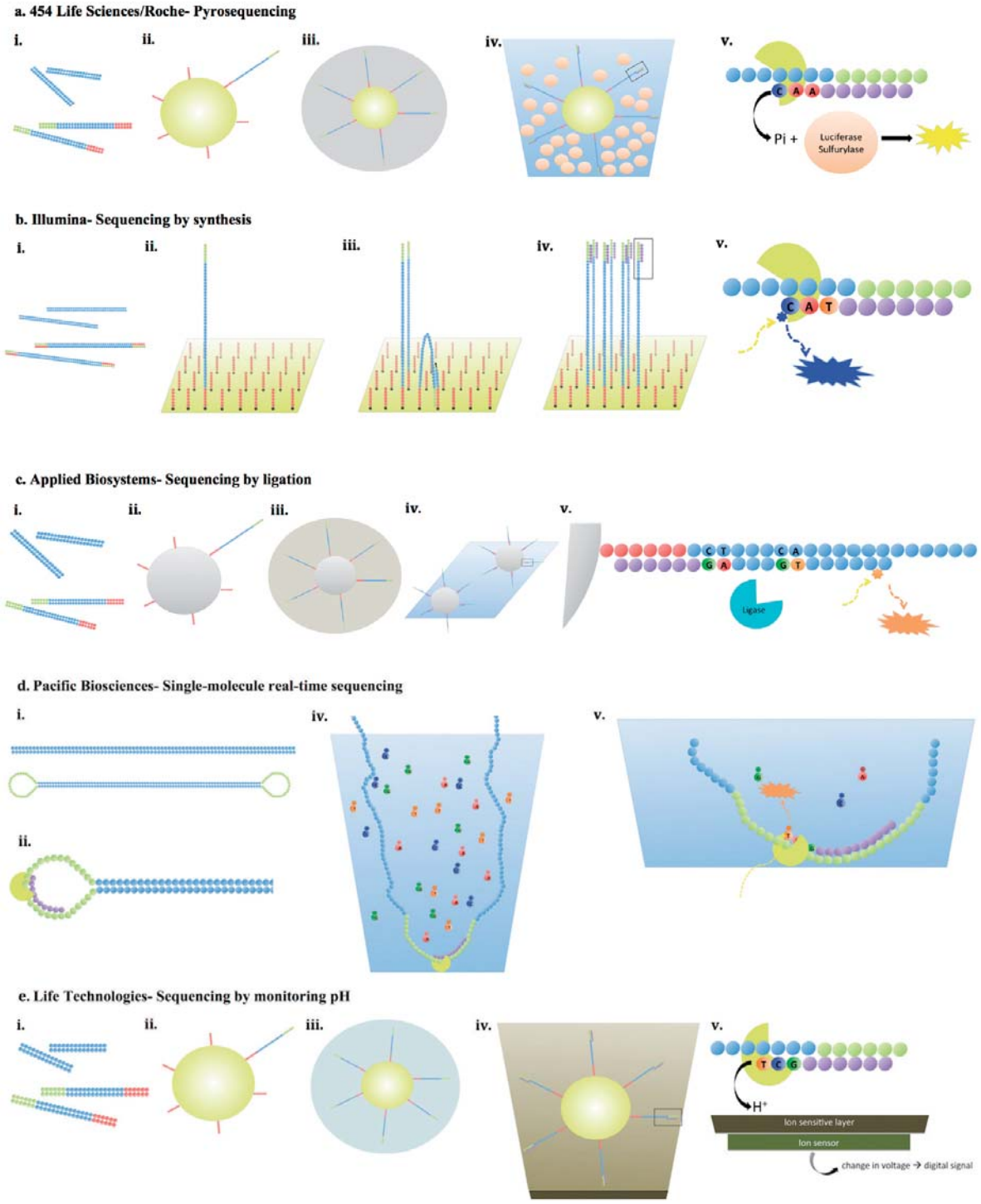



**Figure 9**. Workflow of different NGS technologies. **i.**: DNA fragmentation and ligation of adaptors, **ii.**: Immobilisation to a solid surface, **iii.**: Amplification, **iv.**: Sequencing reaction, **V.**: Signal detection [Nguyen and Burnett 2014].

The general principle of NGS is to generate DNA fragments of a desired size range either by taking PCR products or by fragmenting the DNA using an enzyme mix or a shearing method such as sonication. After size selection, adapters are ligated to the DNA fragments to introduce the primer binding sites for sequencing. The resulting adapter containing fragments are called library and can be generated with or without PCR steps depending on the used chemistry. Furthermore, special adapters including barcode sequences can be used to distinguish different libraries from each other, which make it possible to pool libraries to

optimally use full sequencing capacity of a unit on the sequencer [reviewed in Head *et al*. 2014 and van Dijk *et al*. 2014a]. NGS platforms are divided into two main groups, the so called second-generation sequencing platforms, where the signal is based on many DNA molecules derived from clonal amplification of one library fragment and third-generation sequencing platforms, which are based on single molecule sequencing (Figure 9) [Schadt *et al*. 2010]. After sequencing the signals produced by the sequencer are converted into nucleotide bases with associated quality scores generating short sequences called reads, which are saved in FASTQ files. These reads are subsequently either used for a *de novo* assembly if no reference sequence for the particular genome exists or, as in the case of the human genome, aligned to an already existing reference sequence generating a BAM file. After refinement steps like marking or filtering duplicate reads and realignment around putative small insertions and deletions (INDELs), the files are ready for variant calling. This process generates Variant Call Format (VCF) files, which list called differences to the reference sequence as sequence variants [reviewed in Oliver *et al*. 2015].

The available second-generation sequencing platforms differ on throughput, read length, and sequencing approach (Figure 10, Table 5). Sequencing on the out-dated platform SOLiD (Applied Biosystems, Figure 9c) is based on ligation of fluorescently labelled octomers, whereas on all the other platforms it is based on incorporation of single nucleotides by a polymerase. Accordingly, 454 (Roche, Figure 9a) and IonTorrent (Life Technologies, Figure 9e) offer one of the four different nucleotides at once and the signal generated by its incorporation is recorded, chemiluminescent signal in the presence of released pyrophosphate and changes in pH for 454 and IonTorrent, respectively. In contrast, Illumina (Figure 9b) has differently fluorescently labelled nucleotides with reversible chain terminators. Thus, Illumina allows only the incorporation of one nucleotide at once, whereas in the platforms of 454 and IonTorrent stretches of homopolymers give only one signal. Accordingly, due to variable signal strength per incorporated nucleotide, the accuracy of identifying the number of identical nucleotides decreases with increasing length of a homopolymer leading to high numbers of false positive INDEL calls in 454 and IonTorrent [reviewed in Nguyen and Burnett 2014 and Voelkerding *et al*. 2009]. Illumina, which does not have this issue as well as offers paired-end sequencing and the highest available throughput, is the current market leader [van Dijk *et al*. 2014b].

**Table 5**. Summary of the five major NGS platform families [Hodkinson and Grice 2015].

| *Platform Family* | *Clonal Amplification* | *Chemistry* | *Highest Average Read Length* |
|---|---|---|---|
| 454 | Emulsion PCR | Pyrosequencing (seq-by-synthesis) | 700 bp (paired-end sequencing available) |
| Illumina | Bridge amplification | Reversible dye terminator (seq-by-synthesis) | 300 bp (overlapping paired-end sequencing available) |
| SOLiD | Emulsion PCR | Oligonucleotide 8-mer chained ligation (seq-by-ligation) | 75 bp (paired-end sequencing available) |
| Ion Torrent | Emulsion PCR | Proton detection (seq-by-synthesis) | 400 bp (bidirectional sequencing available) |
| PacBio | N/A (single molecule) | Phospholinked fluorescent nucleotides (seq-by-synthesis) | 8,500 bp |

The average read length is given for the platform/chemistry combination in each family that provides the longest reads.

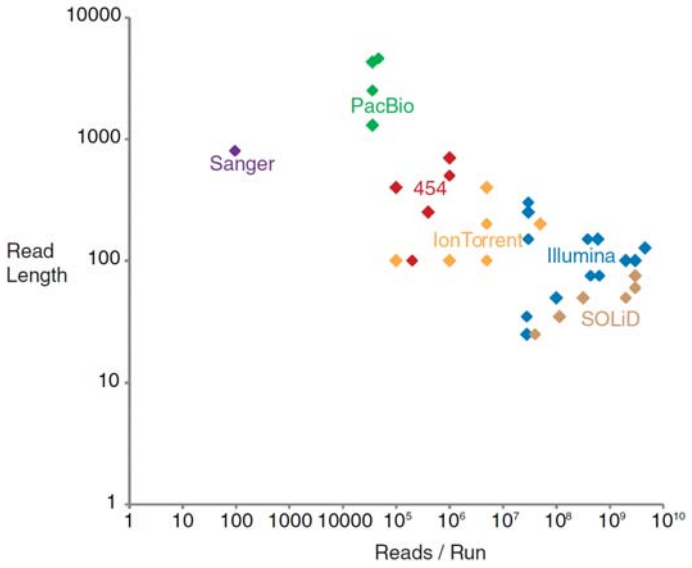



**Figure 10**. Sequencing space based on read length (in bases) and number of reads per run. Points represent official platform/chemistry combination releases and are colour-coded based on the platform family [Hodkinson and Grice 2015].

The advantage of third-generation sequencing (Figure 9d) is not only the absence of clonal amplification, which can be an error source, but also the long reads (>1 kb) and the possibility to detect modification of nucleotides such as methylation by the kinetics of nucleotide incorporation. However, the throughput of third-generation sequencers is currently significantly lower compared to second-generation sequencing platforms (Figure 10). Likewise, the long reads of third-generation sequencers are frequently used in combination with second-generation sequencing data for the closing of gaps in *de novo* assembling, which are most often due to repetitive genomic regions. Currently, there is only one such system on the market, namely the one from PacBio (Figure 9d), which is based on a polymerase, which is attached on the bottom of wells and incorporates fluorescently labelled nucleotides, and a camera that records the signal of the incorporated nucleotides [Schadt *et al*. 2010]. However, a solution called nanopore technology is already under testing before market release [Loman and Watson 2015].

Gene panels with a limited number of genes covered were due to its high coverage and easy interpretable results the first approach that found wide application in a clinical setting. Targets can be enriched by single or multiplex PCR, which is mainly used in smaller panels with a limited number of exons. For larger panels or even exome sequencing, hybridization-based enrichment technologies proved to be more efficient. This can either be by microarrays or, what is more often used, in solution hybridization using biotinylated probes. Such sequencing panels can be custom-designed or commercially available. However, gene panels may need redesign whenever a new gene associated with the corresponding phenotype has been identified. This limitation can be overcome by the application of clinically focussed exome panels or whole-exome sequencing (WES), which

cover all clinically relevant and known exons in the human genome, respectively. However, the exome is not a fixed entity and a subject to change and thus whole-genome sequencing (WGS) is the most comprehensive solution. WGS without any enrichment is less prone to biases and GC-rich regions are better covered. However, since it requires higher amounts of sequencing capacity, WGS is more expensive and lower coverage can be achieved compared to gene panels [reviewed in Xuan *et al*. 2013 and Newman and Black 2014].

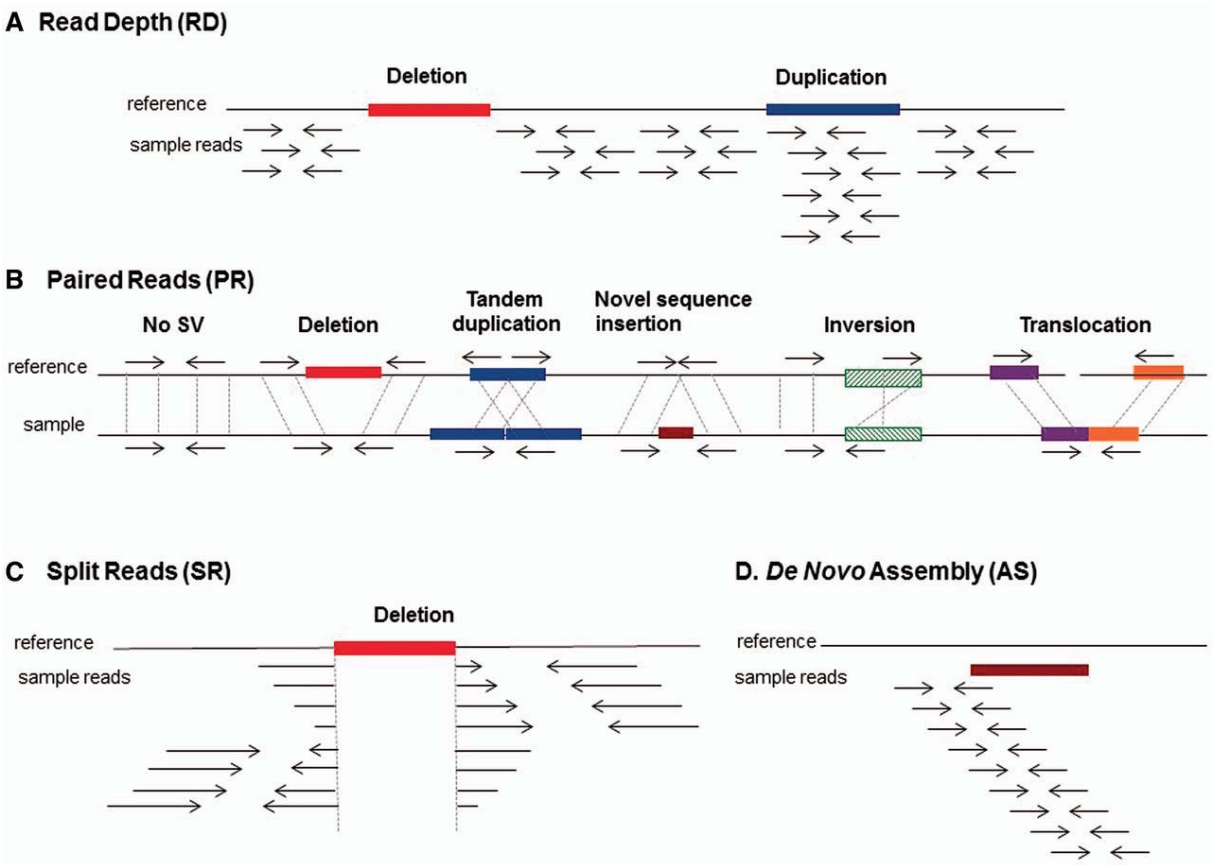



**Figure 11**. Strategies for structural variant (SV) detection. **A**: Read depth. Reads are aligned into the reference genome and when compared to diploid regions they show a reduced number of reads in a deleted region or higher read depth in a duplicated region. **B**: Paired reads. Pairs of sequence reads are mapped into the reference genome (from left to right): (1) no SV, pairs are aligned into correct order, correct orientation, and spanned as expected based on the library's insert size; (2) deletion, the aligned pairs span far apart from that expected based on library insert size; (3) tandem duplication, read pairs are aligned in unexpected order, where expected order means that the leftmost read should be aligned in the forward strand and the rightmost read in the reverse strand; (4) novel sequence insertion, the pairs are aligned closer from that expected based on library insert size; (5) inversion, read pairs are aligned in wrong orientation, both reads align either in forward or reverse strand; and (6) read pairs mapped to different chromosomes. **C**: Split reads. Sequenced reads pointing to the same breakpoint are split at the nucleotide where the breakpoint occurs. The corresponding paired read is properly aligned to the reference genome. **D**: *De novo* assembly. Sample reads from novel sequence insertions are assembled without a reference sequenced genome [Escaramis *et al*. 2015].

NGS allows not only the detection of single nucleotide variants (SNVs) and INDELs but also of CNVs and in contrast to microarrays also copy neutral SVs, especially using WGS data, which have no gaps and biases from enrichment. Different algorithms for the detection of SVs in NGS data are available, which are based on four main strategies (Figure 11). One approach, which allows only the detection of CNVs, is the assessment of changes in read depth (Figure 11A). This approach is the only sequencing-based method to accurately

predict absolute copy number, but is on the other hand hampered by unequal read depth mainly due to differences in GC content and mappability. Furthermore, it is the only approach also applicable to non-WGS data like WES or gene panels, when using algorithms dedicated to this kind of data, which correct for enrichment-based biases. A further approach is the so-called paired reads approach, which considers insert size and orientation of read pairs (Figure 11B). The down side of this approach is that it is prone to false positive calls in repetitive regions due to misalignment. The split reads approach (Figure 11C) realigns unmapped pair mates as putative breakpoint spanning reads by trying to split it to both ends of the deletion. The fourth approach is based on *de novo* assembly (Figure 11D), which, unlike the other approaches, also allows the characterization of novel (non-reverence) sequence insertions. As each of these approaches has its limitations, recent tools for the detection of CNVs in WGS data use a combination of different approaches [reviewed in Xi *et al*. 2012 and Escaramis *et al*. 2015].

There is a wide range of applications for NGS and only a few widely used approaches will be listed here. Next to resequencing for variant calling, DNA sequencing can also be used for *de novo* alignment for species where no reference sequence exists. Chromatin immunoprecipitation followed by NGS (ChIP-seq) enables genome-wide mapping of protein–DNA interactions. Another widely used application is RNA sequencing, which is mostly quantitatively analysed to assess expression levels of genes and isoforms [reviewed in Buermans and den Dunnen 2014].

## 1.3 Aim of the Thesis

Identification of the molecular basis of AD with the current standard approach using Sanger sequencing and multiplex ligation-dependent probe amplification (MLPA) is challenging. Traditional Sanger sequencing is mainly hampered by the high number and large size of the genes known to be associated with AD. Moreover, although the successive application of commonly used exon-by-exon Sanger-sequencing and MLPA is a powerful screening strategy, it misses mutations in non-analyzed gene regions and genes. In addition to these technical limitations, the major scientific problem is that only a limited part of genes mutated in AD is known, preventing gene-targeted analyses.

The aim of this thesis was to contribute to a better understanding of the molecular basis of aortic diseases by applying novel technologies. My working hypothesis was that the molecular basis of AD can be extended by applying appropriate state-of-the-art genome-wide methods to AD cases with unknown aetiology. A large collection of patients with inherited risk for rupture of the aorta or other arteries provided the basis for this thesis. Patients with known sequence variants served as positive controls in evaluation processes.

The knowledge of the molecular basis underlying AD is crucial for (presymptomatic) diagnosis, optimal disease management, and genetic counselling as well as serves as basis for the development of targeted therapeutic strategies. Beyond this thesis, the evaluated methods and developed procedures may also be applied to any other inherited monogenic disorders.

# 2 Results

## 2.1 Published Results

### 2.1.1 Precise Breakpoint Localization of Large Genomic Deletions using PacBio and Illumina Next-Generation Sequencers

Okoniewski MJ*, Meienberg J*, Patrignani A, Szabelska A, Matyas G, Schlapbach R (2013) Precise breakpoint localization of large genomic deletions using PacBio and Illumina next-generation sequencers. Biotechniques 54:98-100.

*Equally contributing first authors

Impact factor: 2.754 (2013)

### 2.1.1.1 Publication

Next-Generation Sequencing | DNA Sequencing Methods | Genomics/ Transcriptomics

BioTechniques Rapid Dispatches

Benchmarks

# Precise breakpoint localization of large genomic deletions using PacBio and Illumina next-generation sequencers

Michal J Okoniewski[1,2*], Janine Meienberg[3,4*], Andrea Patrignani[1], Alicja Szabelska[1,5], Gabor Matyas[3,4,#], Ralph Schlapbach[1,#]

*[1]Functional Genomics Center Zurich, Zurich, Switzerland, [2]Department of Neuroimmunology and Multiple Sclerosis Research, Neurology Clinic, University Hospital, Zurich, Switzerland, [3]Center for Cardiovascular Genetics and Gene Diagnostics, Zurich, Switzerland, [4]Zurich Center of Integrative Human Physiology, University of Zurich, Zurich, Switzerland, [5]Department of Mathematical and Statistical Methods, Poznan University of Life Sciences, Poznan, Poland*

*These authors contributed equally to this work
#These authors jointly directed this work

Keywords: sequencing, large deletions, amplicons, PacBio, Illumina, Sanger

Supplementary material for this article is available at www.BioTechniques.com/article/113992.

Herein we present the applicability of single-molecule (PacBio RS) and second-generation sequencing technology (Illumina) to the characterization of large genomic deletions. By testing samples previously characterized using a Sanger approach, our methods determined that both next-generation sequencing platforms were able to identify the position of deletion breakpoints. Our results point out various advantages of next-generation sequencing platforms when characterizing genomic deletions; however, special attention must be dedicated to identical sequences flanking the breakpoints, such as poly(N) motifs.

The PacBio RS next-generation sequencing (NGS) technology (Pacific Biosciences, Menlo Park, CA, USA) has not only the potential to identify modified bases and thereby characterize methylation patterns (1,2), but it also provides previously unprecedented sequencing read lengths (>2kb), making it useful for quickly improving existing genome assemblies (3). In this study, we used the advantage of such long reads for the characterization of large deletions previously identified by multiplex ligation-dependent probe amplification (MLPA) and microarray analyses. Using traditional Sanger sequencing to characterize large deletions is time-consuming and work-intensive (4,5), increasing the need for effective breakpoint localization. Indeed, for Sanger sequencing a large fragment (2–10kb) containing the breakpoints has to be amplified by long-range PCR (LR-PCR) and subsequently sequenced in order to identify exact breakpoint positions. Furthermore, since Sanger sequencing permits only ~600 bp to be sequenced using one primer, several sets of internal primers are required for large LR-PCR products.

In contrast, NGS may offer simplified sequencing in such cases. Herein, we tested this possibility by using not only the long reads of the PacBio platform, but also the short reads of a second-generation sequencing technology (HiSeq 2000, Illumina, San Diego, CA, USA). Illumina offers stable lengths of short reads (100 bp in this case) with errors most likely to be grouped at the ends of reads (6,7); the PacBio RS reads from this study had a mean length of 2459 bp and random distribution of errors affecting 10–15% of nucleotides. In addition, only a few dedicated computational techniques are available for the characterization of large deletions by NGS (8), making data analysis a challenge.

The three DNA samples used in this study harbor previously characterized large hemizygous deletions. Deletions in sample 44 and sample 70 (of length 26,887 bp and 302,580 bp, respectively) affect the *FBN1* gene in patients with Marfan syndrome (4); a deletion in sample 53B has a size of 3,408,306 bp and comprises the entire *COL3A1* gene in a patient with Ehlers-Danlos syndrome vascular type (5). Accordingly, ~6.5–8.5 kb LR-PCR products were amplified using the Expand Long Template PCR System (Roche Diagnostics, Rotkreuz, Switzerland) as described previously (4,5) and purified by means of QIAquick PCR Purification Kit (Qiagen, Hilden, Germany).

SMRTbell libraries were prepared using the PacBio C2 chemistry (3–10 kb) DNA preparation kit (Part no. 001–540–726,

## Method summary:

We present the applicability of single-molecule (PacBio RS) and second-generation sequencing technology (Illumina) to the characterization of large genomic deletions. Both next-generation sequencing platforms were able to identify the position of deletion breakpoints previously identified by Sanger sequencing.

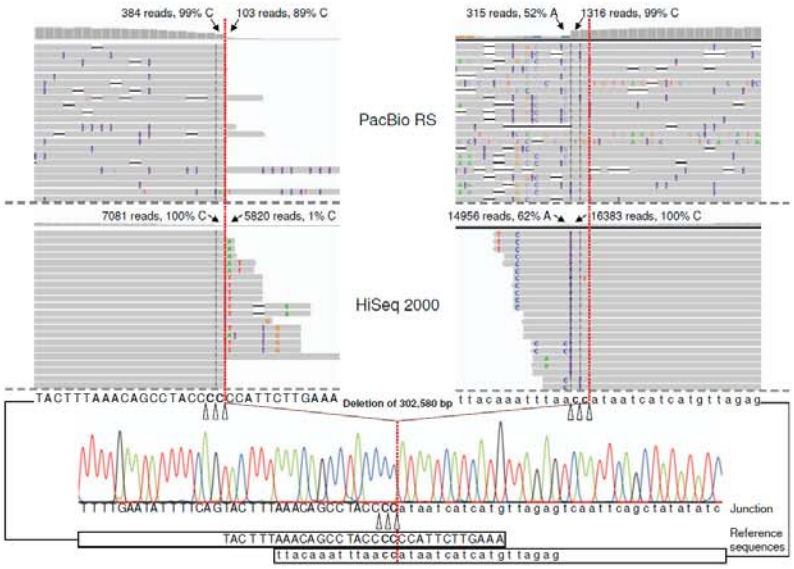



**Figure 1. Both ends of the 302,580-bp deletion on chromosome 15 in sample 70 displayed in the Integrative Genomics Viewer.** Aligned reads are displayed as gray bars/arrows; letters indicate mismatched bases; purple vertical dashes indicate insertions; and black horizontal lines indicate deletions. The reads generated by PacBio RS (upper panel) and Illumina HiSeq 2000 (lower panel) were sorted by aligned position, base, and mapping quality, and compared with the results of Sanger sequencing (bottom). Sections of 22 reads are shown. The top tracks show a much clearer decrease in the PacBio reads, whereas the Illumina reads show a clearer increase in mismatches at the sites of breakpoints. Uppercase letters represent the sequence in the region of the start point of the deletion; lowercase letters represent the sequence in the region of the deletion end point. Due to identical sequences at the sites of breakpoints, the break and re-joining could have occurred at three positions, as indicated by open triangles. The dotted red line marks the most telomeric position of the possible breakpoints. Note that the total read counts (reads) and the percentage of reference bases (%) are given for the positions flanking the site where the coverage (gray bars) starts to lower. For more details, see Tables 1 and 2.

Pacific Biosciences) as well as 5µg purified amplicons without fragmentation. Libraries were subsequently sequenced on the PacBio RS using one SMRT cell per sample and taking two movies of 45 min each. The reads have been mapped with the BLASR mapper (9), which is supplied in the SMRT Portal software suite (Pacific Biosciences) and applies therefore as a standard mapper for PacBio reads. The same amplicons were sequenced on the HiSeq 2000 sequencer using Illumina's TruSeq DNA Sample Preparation v2 protocol with 1 µg input material and 100+100 bp pair-end reads. The reads were mapped using the standard mapper, Bowtie (10). For both NGS platforms, the mappers have been used with default parameters. Respective sequences are available in the NCBI Sequence Read Archive (object ID: ERP002092).

For the PacBio data, the read coverage in the SMRT Portal software suite resulted in a clear drop of read depth in the deleted region (see Supplementary Figure S5), which was subsequently confirmed by zooming in on the breakpoint regions by means of the Integrative Genomics Viewer (IGV) (11) (Figure 1 and Tables 1 and 2 for sample 70; see Supplementary Figures S1–S2 as well as Supplementary Tables S1–S4 for samples 44 and 53B, respectively). Respective Illumina data displayed in IGV showed more gradually sinking patterns at the expected deletion ends; the site of breakpoints in these data was identified by an increase in mismatches (Figure 1, Tables 1 and 2, Supplementary Figures S1–S2, and Supplementary Tables S1–S4). This may be due to the fact that the mappers typically allow several mismatches, meaning that many of the short Illumina reads could be mapped over the breakpoints. In contrast, PacBio RS data show a number of reads spanning over the deletion, which have not been mapped by the SMRT Portal aligner to the standard reference due to the high number of mismatches. The read depth of both platforms is more than sufficient to find the breakpoint; tests using 1/2 or 1/3 of reads per sample also produced satisfactory results (data not shown).

An additional difficulty may be identical sequences on both sides of the deletion, a common phenomenon that has already been described for different genes (12–14).

**Table 1. Read depth and percentage of the wild-type allele in the region flanking the breakpoint at the start site of the deletion in sample 70.**

| Location | Not deleted | | | | | Deleted | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Wild-type Sequence | A | C | C | **C** | **C** | C | C | A | T | T |
| PacBio RS | 614 (100%) | 531 (100%) | 508 (100%) | 492 (100%) | 384 (99%) | 103 (89%) | 66 (88%) | 66 (98%) | 61 (95%) | 29* (86%) |
| HiSeq 2000 | 12699 (100%) | 11679 (100%) | 9545 (100%) | 8278 (100%) | 7081 (100%) | 5820 (1%) | 4313 (48%) | 3224 (100%) | 1678 (99%) | 1536** (98%) |

*No mapped reads 243 bases after the most telomeric breakpoint (read depth = 0). ** No mapped reads 117 bases after the most telomeric breakpoint (read depth = 0). Bold letters indicate identical bases at the sites of the breakpoint, which can be either up- or downstream of the breakpoint. The red dotted line indicates the most telomeric position of the three possible breakpoints. This is also the point where the read depth drops and the number of mismatches increases (see Figure 1).

**Table 2. Read depth and percentage of the wild-type allele in the region flanking the breakpoint at the end of the deletion in sample 70.**

| Location | Deleted | | | | | Not deleted | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Wild-type Sequence | t | t | t | a | a | c | c | a | t | a |
| PacBio RS | 231* (8%) | 242 (10%) | 251 (94%) | 215 (95%) | 315 (52%) | 1316 (99%) | 1359 (99%) | 1411 (99%) | 1465 (99%) | 1547 (100%) |
| HiSeq 2000 | 5408** (100%) | 8762 (87%) | 10427 (83%) | 13679 (92%) | 14956 (62%) | 16383 (100%) | 17735 (96%) | 18658 (100%) | 19512 (100%) | 20549 (100%) |

*No mapped reads 210 bases before the most telomeric breakpoint (read depth = 0). **No mapped reads 16 bases before the most telomeric breakpoint (read depth = 0). Bold letters indicate identical bases at the sites of breakpoints, which can be either up- or downstream of the breakpoint. The red dotted line indicates the most telomeric position of the possible breakpoints. The most centromeric breakpoint, where the read depth drops and the number of mismatches increases, is indicated by a black bold line (see Figure 1).

| PacBio | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | n=0 | n=1 | n=2 | n=3 | n=4 | n=5 | n=6 | n=7 | n=8 | type II error |
| Flank 5 | 56 | 28 | 191 | 941 | 2132 | 348 | 37 | 8 | 0 | 3e-141 |
| *Flank 10* | 0 | 1 | 53 | 329 | 820 | 124 | 12 | 3 | 0 | 1e-46 |
| Flank 20 | 0 | 0 | 5 | 56 | 127 | 19 | 2 | 1 | 0 | 5e-06 |
| **Illumina** | | | | | | | | | | |
| | n=0 | n=1 | n=2 | n=3 | n=4 | n=5 | n=6 | n=7 | n=8 | type II error |
| Flank 5 | 2 | 3 | 0 | 29 | 68817 | 3 | 0 | 0 | 0 | <2e-308 |
| *Flank 10* | 0 | 0 | 0 | 20 | 58769 | 0 | 0 | 0 | 0 | <2e-308 |
| Flank 20 | 0 | 0 | 0 | 13 | 43258 | 0 | 0 | 0 | 0 | <2e-308 |

TCAGTACTTT*AAACA*GCCTA(C)nATAAT*CATCA*TGTTAGAGTC

**Figure 2. Counts of exact matches for different lengths of a poly(C) motif (red) at the site of deletion breakpoints in sample 70 with 5-, 10-, and 20-nucleotide flanking sequences for both PacBio and Illumina reads.** The count indicates that the true sequence includes 4 × C (n = 4) (see Figure 1). Corresponding type II errors were calculated using the R script provided in the Supplementary Materials.

In particular, this could be observed in all three deletions presented in this study ("CC" in samples 53B and 70 and "GC" in sample 44). In order to find the precise sequence of poly(N) motifs (tandemly repeated nucleotides) at the sites of break and rejoining, we have developed an AWK script to count matches at the sites of suspected deletion breakpoints (see Supplementary Materials). This counting was performed with perfect matches only, resulting in the data depicted in Figure 2 (sample 70) and Supplementary Figures S3 and S4 (samples 44 and 53B, respectively). When a single nucleotide (or pair, in the case of GC) has a fixed probability of being misinterpreted, it can be assumed without loss of generality that the distribution of the occurrences of specific motifs follows the Poisson distribution. The hypothesis that the maximum number of counts represents the appropriate motif has been tested. For PacBio RS reads in sample 70, the probabilities of wrongly accepting the null hypothesis are far below the 0.01 level of significance ($p = 1.5 \times 10^{-23}$, $p = 1.06 \times 10^{-46}$, and $p = 3.2 \times 10^{-141}$ in the cases of 20, 10, and 5 flanking bases, respectively) (Figure 2). In the case of Illumina, due to the high number of reads, error levels are so low that they go below that afforded by the small-number precision in the R language. For details on the calculations, see the R script in the Supplementary Materials. The script can be used on any FASTA or FASTQ data and checks the statistical power at a given significance level regardless of the platform.

As shown by this study, the determination of deletion breakpoints can be done with data obtained from both NGS platforms. However, whereas the long reads of PacBio RS showed a sharp decrease in read depth, Illumina short reads exhibited an increase in mismatches related to the position of the breakpoints. Sample preparation costs are comparable for PacBio and Illumina platforms. However, sequencing using PacBio RS can be done within a working day, while Illumina's system, even the smaller MiSeq version, requires more time. Both platforms are suitable for precise breakpoint localization, provide an alternative procedure for the characterization of large deletions, and require fewer resources and less time than traditional Sanger sequencing.

## Acknowledgments

We are grateful to Yu-Chih Tsai, Jonas Korlach, and Stephen W. Turner for discussions on PacBio technology and data analysis. This work was supported by the FGCZ, as well as grants from the COFRA Foundation (to G.M.), Gottfried & Julia Bangerter-Rhyner-Stiftung (to G.M.), Jubiläumsstiftung Swiss Life (to G.M.), Foundation for People with Rare Diseases (to J.M. and G.M.), Clinical Research Priority Program (CRPP/KFSP-MS) of University of Zurich (to M.O.), and Sciex.ch (no. 11.182 to A.S. and M.O.).

## Competing interests

The authors declare no competing interests.

## References

1. **Clark, T.A., I.A. Murray, R.D. Morgan, A.O. Kislyuk, K.E. Spittle, M. Boitano, A. Fomenkov, R.J. Roberts, and J. Korlach.** 2012. Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. Nucleic Acids Res. *40*:e29.
2. **Murray, I.A., T.A. Clark, R.D. Morgan, M. Boitano, B.P. Anton, K. Luong, A. Fomenkov, S.W. Turner, et al.** 2012. The methylomes of six bacteria. Nucleic Acids Res. *40*:11450-11462.
3. **Zhang, X., K.W. Davenport, W. Gu, H.E. Daligault, A.C. Munk, H. Tashima, K. Reitenga, L.D. Green, and C.S. Han.** 2012. Improving genome assemblies by sequencing PCR products with PacBio. BioTechniques *53*:61-62.
4. **Mátyás, G., S. Alonso, A. Patrignani, M. Marti, E. Arnold, I. Magyar, C. Henggeler, T. Carrel, et al.** 2007. Large genomic fibrillin-1 (FBN1) gene deletions provide evidence for true haploinsufficiency in Marfan syndrome. Hum. Genet. *122*:23-32.
5. **Meienberg, J., M. Rohrbach, S. Neuenschwander, K. Spanaus, C. Giunta, S. Alonso, E. Arnold, C. Henggeler, et al.** 2010. Hemizygous deletion of COL3A1, COL5A2, and MSTN causes a complex phenotype with aortic dissection: a lesson for and from true haploinsufficiency. Eur. J. Hum. Genet. *18*:1315-1321.
6. **Kozarewa, I., Z. Ning, M.A. Quail, M.J. Sanders, M. Berriman, and D.J. Turner.** 2009. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. Nat. Methods *6*:291-295.
7. **McElroy, K.E., F. Luciani, and T. Thomas.** 2012. GemSIM: general, error-model based simulator of next-generation sequencing data. BMC Genomics *13*:74.
8. **Ye, K., M.H. Schulz, Q. Long, R. Apweiler, and Z. Ning.** 2009. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics *25*:2865-2871.
9. **Chaisson, M.J. and G. Tesler.** 2012. Mapping single molecule sequencing reads using Basic Local Alignment with Successive Refinement (BLASR): Theory and application. BMC Bioinformatics *13*:238.
10. **Langmead, B.** 2010. Aligning short sequencing reads with Bowtie. Curr. Protoc. Bioinformatics *Chapter 11*:Unit 11.7.
11. **Thorvaldsdottir, H., J.T. Robinson, and J.P. Mesirov.** 2012. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform [Epub ahead of print].
12. **Giacalone, J.P. and U. Francke.** 1992. Common sequence motifs at the rearrangement sites of a constitutional X/autosome translocation and associated deletion. Am. J. Hum. Genet. *50*:725-741.
13. **Otto, E., R. Betz, C. Rensing, S. Schatzle, T. Kuntzen, T. Vetsi, A. Imm, and F. Hildebrandt.** 2000. A deletion distinct from the classical homologous recombination of juvenile nephronophthisis type 1 (NPH1) allows exact molecular definition of deletion breakpoints. Hum. Mutat. *16*:211-223.
14. **Liu, H.X., L. Cartegni, M.Q. Zhang, and A.R. Krainer.** 2001. A mechanism for exon skipping caused by nonsense or missense mutations in BRCA1 and other genes. Nat. Genet. *27*:55-58.

Received 26 October 2012; accepted 8 January 2013.

Address correspondence to Michal J. Okoniewski, FGCZ, Winterthurerstrasse 190, 8057 Zurich, Switzerland. Email: michal.okoniewski@fgcz.ethz.ch

Next-Generation Sequencing | DNA Sequencing Methods | Genomics/ Transcriptomics

BioTechniques Rapid Dispatches

**Supplementary Material for:**

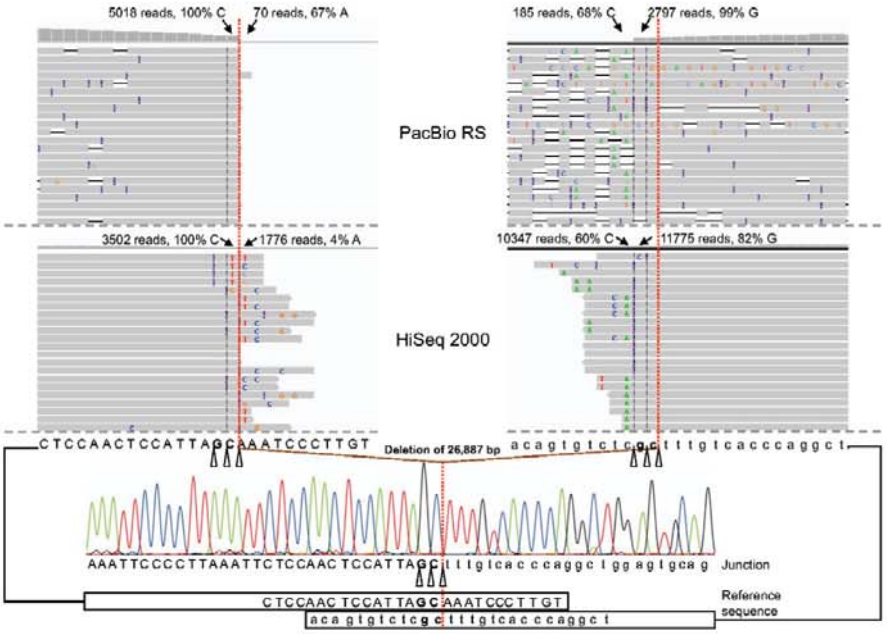
# Precise breakpoint localization of large genomic deletions using PacBio and Illumina next-generation sequencers

Michal J Okoniewski[1,2*], Janine Meienberg[3,4*], Andrea Patrignani[1], Alicja Szabelska[1,5], Gabor Matyas[3,4,#], Ralph Schlapbach[1,#]

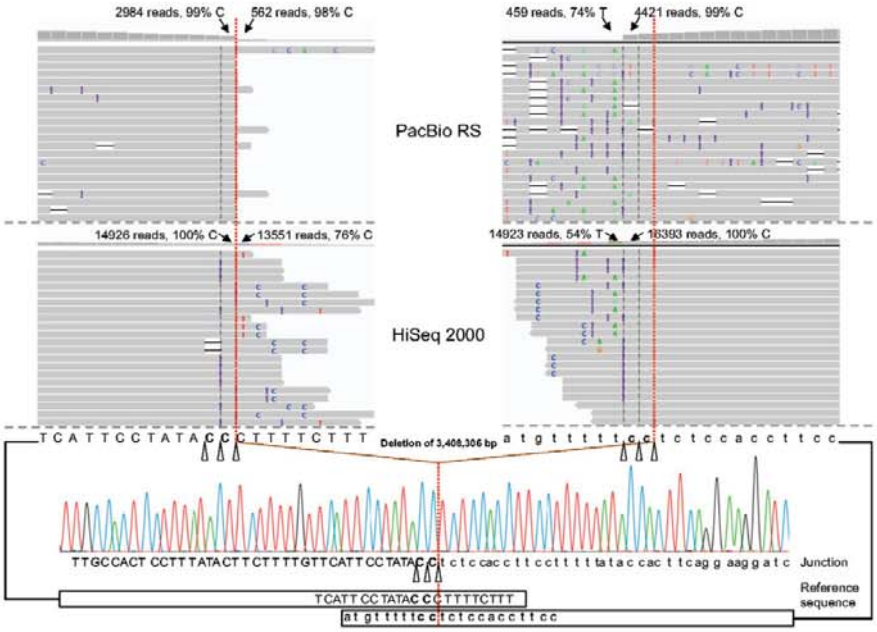
*[1]Functional Genomics Center Zurich, Zurich, Switzerland, [2]Department of Neuroimmunology and Multiple Sclerosis Research, Neurology Clinic, University Hospital, Zurich, Switzerland, [3]Center for Cardiovascular Genetics and Gene Diagnostics, Zurich, Switzerland, [4]Zurich Center of Integrative Human Physiology, University of Zurich, Zurich, Switzerland, [5]Department of Mathematical and Statistical Methods, Poznan University of Life Sciences, Poznan, Poland*

*These authors contributed equally to this work
#These authors jointly directed this work

Keywords: sequencing, large deletions, amplicons, PacBio, Illumina, Sanger




**Supplementary Figure S1. Both ends of the 26,887bp deletion on chromosome 15 in sample 44 displayed in the Integrative Genomics Viewer.** The reads generated by the PacBio RS (upper panel) and the Illumina HiSeq 2000 (lower panel) were sorted by aligned position, base, and mapping quality, and compared with the results of Sanger sequencing (bottom). Sections of 22 reads are shown. Aligned reads are displayed as gray bars/arrows, letters indicate mismatched bases, purple vertical dashes insertions, and black horizontal lines deletions. Note that the total read counts (reads) and the percentage of reference bases (%) are given for the positions flanking the site where the coverage (gray bars) starts to lower. Uppercase letters represent the sequence in the region of the start point of the deletion; lowercase letters represent the sequence in the region of the deletion end point. Due to identical sequences at the site of breakpoints, the break and rejoining could have occurred at three positions, as indicated by open triangles. The dotted red line marks the most telomeric position of the possible breakpoints. For more details, see Supplementary Tables S1 and S2.




**Supplementary Figure S2. Both ends of the 3,408,306bp deletion on chromosome 2 in sample 53B displayed in the Integrative Genomics Viewer.** The reads generated by the PacBio RS (upper panel) and the Illumina HiSeq 2000 (lower panel) were sorted by aligned position, base, and mapping quality, and compared with the results of Sanger sequencing (bottom). Sections of 22 reads are shown. Aligned reads are displayed as gray bars/arrows, letters indicate mismatched bases, purple vertical dashes insertions, and black horizontal lines deletions. Note that the total read counts (reads) and the percentage of reference bases (%) are given for the positions flanking the site where the coverage (gray bars) starts to lower. Uppercase letters represent the sequence in the region of the start point of the deletion; lowercase letters represent the sequence in the region of the deletion end point. Due to identical sequences at the site of breakpoints, the break and rejoining could have occurred at three positions, as indicated by open triangles. The dotted red line marks the most telomeric position of the possible breakpoints. For more details, see Supplementary Tables S3 and S4.

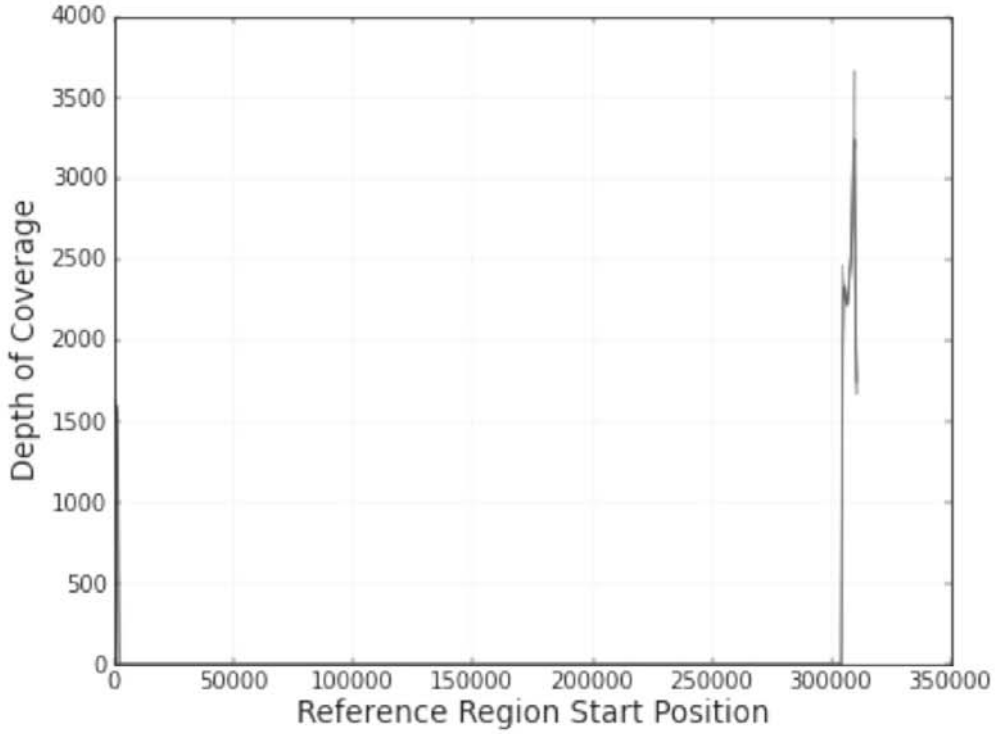
| PacBio | | | | |
|---|---|---|---|---|
| | n=0 | n=1 | n=2 | type II error |
| Flank 5 | 885 | 2270 | 11 | <2e-308 |
| *Flank 10* | 0 | 649 | 4 | 6e-61 |
| Flank 20 | 0 | 74 | 1 | 4e-06 |
| **Illumina** | | | | |
| | n=0 | n=1 | n=2 | type II error |
| Flank 5 | 116 | 52218 | 0 | <2e-308 |
| *Flank 10* | 0 | 44249 | 0 | <2e-308 |
| Flank 20 | 0 | 32717 | 0 | <2e-308 |

TAAATTCTCC*AACTC*CATTA(GC)nTTTGT*CACCC*AGGCTGGAGT

**Supplementary Figure S3. Counts of exact matches for different lengths of a GC-motif (red) at the site of deletion breakpoints in sample 44 with 5-, 10-, and 20-nucleotide flanking sequences for both PacBio and Illumina reads (denoted by different shades of green).** Counts indicate that the true sequence includes 1 × GC ($n = 1$) (see Supplementary Figure S1). Corresponding type II errors were calculated using the R script provided in the Supplementary Materials.

| PacBio | | | | | | |
|---|---|---|---|---|---|---|
| | n=0 | n=1 | n=2 | n=3 | n=4 | type II error |
| Flank 5 | 200 | 888 | 2904 | 244 | 12 | 6e-280 |
| *Flank 10* | 19 | 253 | 1026 | 61 | 2 | 3e-84 |
| Flank 20 | 3 | 41 | 186 | 12 | 1 | 6e-13 |
| **Illumina** | | | | | | |
| | n=0 | n=1 | n=2 | n=3 | n=4 | type II error |
| Flank 5 | 14 | 1 | 53212 | 0 | 0 | <2e-308 |
| *Flank 10* | 0 | 0 | 45448 | 0 | 0 | <2e-308 |
| Flank 20 | 0 | 0 | 32554 | 0 | 0 | <2e-308 |

**Supplementary Figure S4. Counts of exact matches for different lengths of a poly(C) motif (red) at the site of deletion breakpoints in sample 53B with 5-, 10-, and 20-nucleotide flanking sequences for both PacBio and Illumina reads (denoted by different shades of green).** Counts indicate that the true sequence includes 2 × C ($n = 2$) (see Supplementary Figure 2). Corresponding type II errors were calculated using the R script provided in the Supplementary Materials.




**Supplementary Figure S5. Coverage plot from the PacBio software, SMRTportal.** Plot is shown for sample 70 and is based upon a standard human genome reference.

**Supplementary Table S1. Read Depth and Percentage of Wild-type Allele in the Region Flanking the Breakpoint at the Start Site of the Deletion in Sample 44.**

| Location | Not Deleted | | | | | Deleted | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Wild-type Sequence | T | T | A | **G** | **C** | A | A | A | T | C |
| PacBio RS | 7226 (99%) | 6738 (99%) | 6144 (99%) | 5621 (99%) | 5018 (100%) | 70 (67%) | 35 (40%) | 43 (37%) | 150 (96%) | 39* (79%) |
| HiSeq 2000 | 9403 (100%) | 7894 (100%) | 6118 (100%) | 4626 (100%) | 3502 (100%) | 1776 (4%) | 845 (2%) | 92 (20%) | Fifteen (53%) | 6** (67%) |

*No mapped reads 468 bases after the most telomeric breakpoint (read depth = 0).
**No mapped reads 14 bases after the most telomeric breakpoint (read depth = 0).
Bold letters indicate identical bases at the site of the breakpoint, which can be either up- or downstream of the breakpoint. The red dotted line indicates the most telomeric position of the three possible breakpoints. This is also the point where the read depth drops and the number of mismatches increases (see Supplementary Figure S1).

**Supplementary Table S2. Read Depth and Percentage of Wild-type Allele in the Region Flanking the Breakpoint at the End of the Deletion in Sample 44.**

| Location | Deleted | | | | | Not Deleted | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Wild-type sequence | g | t | c | t | c | **g** | **c** | t | t | t |
| PacBio RS | 77* (34%) | 104 (72%) | 71 (76%) | 132 (72%) | 185 (68%) | 2797 (99%) | 3061 (99%) | 3495 (99%) | 4025 (99%) | 4219 (100%) |
| HiSeq 2000 | 2200** (4%) | 5594 (69%) | 8918 (87%) | 9609 (83%) | 10347 (60%) | 11775 (82%) | 13756 (100%) | 14409 (100%) | 14934 (99%) | 15995 (100%) |

*No mapped reads 368 bases before the most telomeric breakpoint (read depth = 0).
**No mapped reads 39 bases before the most telomeric breakpoint (read depth = 0).
Bold letters indicate identical bases at the site of the breakpoints, which can be either up- or downstream of the breakpoint. The red dotted line indicates the most telomeric position of the possible breakpoints. The most centromeric breakpoint, where the read depth drops and the number of mismatches increases, is indicated by a black bold line (see Supplementary Figure S1).

**Supplementary Table S3. Read Depth and Percentage of Wild-type Allele in the Region Flanking the Breakpoint at the Start Site of the Deletion in Sample 53B.**

| Location | Not deleted | | | | | Deleted | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Wild-type Sequence | A | T | A | **C** | **C** | C | T | T | T | T |
| PacBio RS | 4386 (99%) | 4132 (99%) | 3931 (99%) | 3568 (99%) | 2984 (99%) | 562 (98%) | 493 (96%) | 211 (60%) | 190 (59%) | 192* (92%) |
| HiSeq 2000 | 20179 (100%) | 18897 (100%) | 17702 (100%) | 14635 (100%) | 14926 (100%) | 13551 (76%) | 12489 (70%) | 11602 (71%) | 8538 (99%) | 8470** (60%) |

*No mapped reads 544 bases after the most telomeric breakpoint (read depth = 0).
**No mapped reads 13 bases after the most telomeric breakpoint (read depth = 0).
Bold letters indicate identical bases at the site of the breakpoint, which can be either up- or downstream of the breakpoint. The red dotted line indicates the most telomeric position of the three possible breakpoints. This is also the point where the read depth drops and the number of mismatches increases (see Supplementary Figure S2).

**Supplementary Table S4. Read Depth and Percentage of Wild-type Allele in the Region Flanking the Breakpoint at the End of the Deletion in Sample 53B.**

| Location | Deleted | | | | | Not deleted | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Wild-type Sequence | t | t | t | t | t | **c** | **c** | t | c | t |
| PacBio RS | 223* (81%) | 230 (79%) | 240 (61%) | 331 (91%) | 459 (74%) | 4421 (99%) | 5192 (100%) | 5678 (100%) | 6066 (100%) | 6516 (100%) |
| HiSeq 2000 | 7818** (73%) | 9888 (100%) | 10675 (57%) | 13335 (89%) | 14923 (54%) | 16393 (100%) | 17832 (100%) | 19429 (100%) | 20646 (100%) | 21893 (100%) |

*No mapped reads 411 bases before the most telomeric breakpoint (read depth = 0).
**No mapped reads 11 bases before the most telomeric breakpoint (read depth = 0).
Bold letters indicate identical bases at the site of the breakpoints, which can be either up- or downstream of the breakpoint. The red dotted line indicates the most telomeric position of the possible breakpoints. The most centromeric breakpoint, where the read depth drops and the number of mismatches increases, is indicated by a black bold line (see Supplementary Figure S2).

## AWK Script (Example for sample 70)

```
BEGIN {c8=0; c7=0; c6=0; ;c5=0; c4=0; c3=0; c2=0; c1=0; c0=0; sekw="C"; l=0}
{ if ($0 ~ />/)
  {   if (sekw ~ /TCAGTACTTTAAACAGCCTACCCCCCCCCATAATCATCATGTTAGAGTC/) c8=c8+1;
   if (sekw ~ /TCAGTACTTTAAACAGCCTACCCCCCCATAATCATCATGTTAGAGTC/) c7=c7+1;
   if (sekw ~ /TCAGTACTTTAAACAGCCTACCCCCCATAATCATCATGTTAGAGTC/) c6=c6+1;
   if (sekw ~ /TCAGTACTTTAAACAGCCTACCCCCATAATCATCATGTTAGAGTC/) c5=c5+1;
   if (sekw ~ /TCAGTACTTTAAACAGCCTACCCCATAATCATCATGTTAGAGTC/) c4=c4+1;
   if (sekw ~ /TCAGTACTTTAAACAGCCTACCCATAATCATCATGTTAGAGTC/) c3=c3+1;
   if (sekw ~ /TCAGTACTTTAAACAGCCTACCATAATCATCATGTTAGAGTC/) c2=c2+1;
   if (sekw ~ /TCAGTACTTTAAACAGCCTACATAATCATCATGTTAGAGTC/) c1=c1+1;
   if (sekw ~ /TCAGTACTTTAAACAGCCTAATAATCATCATGTTAGAGTC/) c0=c0+1;
   l=l+length(sekw)
   sekw = "";  }
 else sekw=sekw$0;}
END { print (c0,"",c1,"",c2,"",c3,"",c4,"",c5,"",c6,"",c7,"", c8,"",l)}

   if (sekw ~ /TCAGTACTTTAAACAGCCTACCCCCATAATCATCATGTTAGAGTC/) c5=c5+1;
   if (sekw ~ /TCAGTACTTTAAACAGCCTACCCCATAATCATCATGTTAGAGTC/) c4=c4+1;
   if (sekw ~ /TCAGTACTTTAAACAGCCTACCCATAATCATCATGTTAGAGTC/) c3=c3+1;
   if (sekw ~ /TCAGTACTTTAAACAGCCTACCATAATCATCATGTTAGAGTC/) c2=c2+1;
   if (sekw ~ /TCAGTACTTTAAACAGCCTACATAATCATCATGTTAGAGTC/) c1=c1+1;
   if (sekw ~ /TCAGTACTTTAAACAGCCTAATAATCATCATGTTAGAGTC/) c0=c0+1;
   l=l+length(sekw)
   sekw = "";
  }
 else sekw=sekw$0;
}
END { print (c0,"",c1,"",c2,"",c3,"",c4,"",c5,"",c6,"",c7,"", c8,"",l)}
```

## R Script

```
dane<-read.csv('polyCmotifs.csv', header=T) #loading data with counts of the exact matches with 5,10,20-nucleotide flanking region
a<-0.01 # significance level of the test

z<-qnorm(1-a) # z statistics needed for the calculations of the type II error
n<-rowSums(dane) # total number of the number of
l1<-40
l1<-c(10,20,40) #length of flanking region
p<-0.12 # probability of mismatch
N<-83830 # total number of reads
psum1<-dnbinom(l1,1,p) #estimated probability that there is no error in the sequence, negative binomial distribution is assumed
n.reads1<-N*psum1 # expected number of exact matches with chosen flanking region

#calculation of the type II error (probability that the null hypothesis was wrongly accepted)
beta<-NULL
x<-seq(0,8,by=1) #considered number of deletions

for (i in 1:3)
{   lambda<-as.numeric(colnames(dane)[which(dane[i,]==max(dane[i,]))]) # choice of the motif with maximun counts of the exact matches
  x<-x[-lambda] #obtaining possible alternative hypotheses by exculing the lambda from considered cases

  #calculations of errors has to be divided in 2 cases, when alternative is smaller or higher than null hypothesis
  y<-x[x<lambda]
  b<-1-pnorm((lambda-z*sqrt(lambda/n[i])-y)/sqrt(y/n[i])) # type II error for alternative hypotheses < lambda

  y<-x[x>lambda]
  b<-c(b,pnorm((lambda+z*sqrt(lambda/n[i])-y)/sqrt(y/n[i])))# type II error for alternative hypotheses > lambda

  beta<-c(beta,sum(b))          }

print(beta)
```

### 2.1.1.2 Contribution of Authors

| | |
|---|---|
| Michal Okoniewski | Equally contributing first author; data analysis, writing of the manuscript |
| **Janine Meienberg** | **Equally contributing first author; library preparation, interpretation of data, writing of the manuscript** |
| Andrea Patrignani | Library preparation and sequencing, writing of the manuscript |
| Alicja Szabelska | Statistical calculations, writing of the manuscript |
| Gabor Matyas | Equally contributing senior author; conceptual planning and design of the study, writing and editing of the manuscript |
| Ralph Schlapbach | Equally contributing senior author; initiation of the study, writing of the manuscript |

### 2.1.2 New Insights into the Performance of Human Whole-Exome Capture Platforms

Meienberg J*, Zerjavic K*, Keller I, Okoniewski M, Patrignani A, Ludin K, Xu Z, Steinmann B, Carrel T, Roethlisberger B, Schlapbach R, Bruggmann R, Matyas G (2015) New insights into the performance of human whole-exome capture platforms. Nucleic Acids Res 43:e76.

*Equally contributing first authors

Impact factor: 8.808 (2013)

## 2.1.2.1 Publication

**Nucleic Acids Research Advance Access published March 27, 2015**

*Nucleic Acids Research, 2015* **1**
*doi: 10.1093/nar/gkv216*

# New insights into the performance of human whole-exome capture platforms

**Janine Meienberg[1,†], Katja Zerjavic[1,†], Irene Keller[2], Michal Okoniewski[3,4], Andrea Patrignani[3], Katja Ludin[5], Zhenyu Xu[6], Beat Steinmann[7], Thierry Carrel[8], Benno Röthlisberger[5], Ralph Schlapbach[3], Rémy Bruggmann[9] and Gabor Matyas[1,8,10,*]**

[1]Center for Cardiovascular Genetics and Gene Diagnostics, Foundation for People with Rare Diseases, Schlieren-Zurich CH-8952, Switzerland, [2]Department of Clinical Research, University of Berne, Berne CH-3010, Switzerland, [3]Functional Genomics Center Zurich, Zurich CH-8057, Switzerland, [4]Division of Scientific IT Services, ETH Zurich, Zurich CH-8092, Switzerland, [5]Division of Medical Genetics, Center for Laboratory Medicine, Aarau CH-5001, Switzerland, [6]Sophia Genetics SA, Lausanne CH-1015, Switzerland, [7]Division of Metabolism, University Children's Hospital, Zurich CH-8032, Switzerland, [8]Department of Cardiovascular Surgery, University Hospital, Berne CH-3010, Switzerland, [9]Interfaculty Bioinformatics Unit and Swiss Institute of Bioinformatics, University of Berne, Berne CH-3012, Switzerland and [10]Zurich Center for Integrative Human Physiology, University of Zurich, Zurich CH-8057, Switzerland

Received November 17, 2014; Revised March 2, 2015; Accepted March 3, 2015

## ABSTRACT

**Whole exome sequencing (WES) is increasingly used in research and diagnostics. WES users expect coverage of the entire coding region of known genes as well as sufficient read depth for the covered regions. It is, however, unknown which recent WES platform is most suitable to meet these expectations. We present insights into the performance of the most recent standard exome enrichment platforms from Agilent, NimbleGen and Illumina applied to six different DNA samples by two sequencing vendors per platform. Our results suggest that both Agilent and NimbleGen overall perform better than Illumina and that the high enrichment performance of Agilent is stable among samples and between vendors, whereas NimbleGen is only able to achieve vendor- and sample-specific best exome coverage. Moreover, the recent Agilent platform overall captures more coding exons with sufficient read depth than NimbleGen and Illumina. Due to considerable gaps in effective exome coverage, however, the three platforms cannot capture all known coding exons alone or in combination, requiring improvement. Our data emphasize the importance of evaluation of updated platform versions and suggest that enrichment-free whole genome sequencing can overcome the limitations of WES in sufficiently covering coding exons, especially GC-rich regions, and in characterizing structural variants.**

## INTRODUCTION

As a widely used method in genomic research and gene diagnostics, whole exome sequencing (WES) has the potential both to capture the entire coding region of all known genes including flanking intronic regions and to provide sequence data from these enriched genomic regions with sufficient read depth using a high-throughput DNA sequencing technology (1–6). Without enrichment, whole genome sequencing (WGS) is more comprehensive and allows for the characterization of the entire genome (7–10). As the human exome represents only ~2% of the human genome, but harbours ~85% of all known disease-causing mutations, WES is an increasingly used alternative to WGS (11–16).

Since performance comparisons of three major commercial exome enrichment platforms from Agilent, NimbleGen and Illumina were reported (2–6), new versions of these platforms have been introduced. To date, however, it has not been shown which new platform is superior in performance and most suitable for diagnostic purposes. For this reason and because our preliminary study (Supplementary Figures S1 and S2) revealed that an updated version does not necessarily lead to performance improvements, we assessed the most recent version (v) of each platform (updated end 2013).

Here, from the perspective of WES users, we present a performance evaluation of exome captures from Agilent (SureSelect v5+UTR), NimbleGen (SeqCap v3+UTR) and

*To whom correspondence should be addressed. Tel: +41 43 433 86 86; Fax: +41 43 433 86 85; Email: matyas@genetikzentrum.ch
†These authors contributed equally to the paper as first authors.

Illumina (Nextera Expanded Exome) applied to six human DNA samples extracted from blood, saliva or cultured cells and sequenced by different providers (vendors). One vendor (V1) used all three platforms and applied the same data analysis workflow (e.g. mapping, variant calling), ensuring best comparability. In order to avoid vendor bias and to assess the reproducibility of capture performance, each platform was used by an additional vendor specialized in the respective platform (V2, V3 and V4) as well, resulting in two sets of WES data per platform for all six samples. This study design allowed us to evaluate and compare the platforms not only within the same experimental and bioinformatics setting, similar to studies of previous platform versions (2–6), but also between different settings and among different DNA sources, revealing hitherto unreported performance differences, drawbacks and capabilities. Regarding coverage of the entire coding exome, for two samples we extended the performance evaluation of WES into the area of WGS as well, including data of the most recent HiSeq X Ten system.

## MATERIALS AND METHODS

### Preliminary study

Exomes of eight DNA samples (including all six samples of this study, Table 1) were enriched using previous capture platforms from Agilent (SureSelect Human All Exon kit v4+UTR) and NimbleGen (SeqCap EZ Human Exome v3) according to the manufacturer's instructions. Subsequently, prepared libraries were sequenced with 2 × 100 bp paired-end reads on a HiSeq 2000 sequencer (Illumina) at the Functional Genomics Center Zurich, including data analysis, which revealed distinct differences in the performance of these previous enrichment platforms (Supplementary Figures S1 and S2).

### Samples

DNA samples of six unrelated individuals, from which informed consent was obtained, were selected for this study. In each of these samples, Sanger sequencing and, if available, multiplex ligation-dependent probe amplification (MLPA) was previously performed for at least a subset of genes listed in Supplementary Table S1 (17–20).

Genomic DNA was extracted from EDTA-anticoagulated peripheral whole blood samples, saliva and cells cultured from aortic walls or skin biopsies (fibroblasts) using either QIAamp DNA Mini kit (Qiagen) or Chemagic Magnetic Separation Module I (Chemagen, Perkin Elmer) according to the manufacturer's instructions. Subsequently, some of the Chemagic extracted DNA samples were purified using QIAamp DNA Mini kit as well (Table 1). DNAs were quantified by OD measurements (NanoDrop 2000, Thermo Scientific) and used for exome enrichment according to the platform's standard requirements, i.e. 3 μg for Agilent, 1 μg for NimbleGen and 50 ng for Illumina.

### Exome enrichment and high-throughput sequencing

Exome enrichment was performed using the most recent versions (available by the end of 2013) of commercial sequence capture kits from Agilent (SureSelect Human All Exon kit v5+UTR), NimbleGen (SeqCap EZ Exome (v3) +UTR) and Illumina (Nextera Rapid Capture Expanded Exome). The captured libraries were sequenced using a HiSeq 2000/2500 sequencer (Illumina) and 2 × 100 bp paired-end sequencing according to the manufacturers' recommendations. As the size of the designed target regions differs among platforms (Supplementary Table S2), the total expected read number was adjusted by the vendors to obtain 100× coverage in all cases (i.e. sequencing at 100× was requested for each sample). In addition to or instead of this comparable coverage for each platform, the usage of comparable number of reads was not pursued as this would favour the platform with the smallest exome design (Illumina).

Library preparation, sequence capture and high-throughput sequencing was performed by four different sequencing service providers (vendors V1–V4) according to their standard workflow. One vendor (V1) sequenced all six genomic DNA samples using all three exome enrichment platforms, whereas the three other vendors sequenced the six DNAs using only one platform in which they were specialized (V2: Agilent; V3: NimbleGen; V4: Illumina; Table 1). To extend our evaluation of the effect of DNA sources, 24 additional DNA samples extracted from blood (18 samples) or saliva (6 samples) were sequenced by V2 at 60× using Agilent capture (SureSelect v5) (Supplementary Figure S3).

In addition to the three updated standard exome enrichment platforms, the targeted (focused) enrichment of ∼7600 clinically relevant genes was exemplified for sample 7739 and two additional DNA samples using the Accuracy and Content Enhanced (ACE) clinical exome platform of Personalis with ∼4 μg DNA sequenced at 60× on a HiSeq2500 system (ACEv2, www.personalis.com). Moreover, for two out of the six DNA samples (7344 and 7739) WGS was also performed by three vendors (V1, V3 and V4) using Illumina's TruSeq DNA PCR-Free Sample Preparation Kit on a HiSeq 2000/2500 sequencer (Table 1). V1 performed WGS at 60× coverage using 4 μg DNA (sample 7739), whereas V3 (samples 7344 and 7739) and V4 (sample 7739) carried out sequencing at 30× using 2 and 4 μg DNA, respectively. Furthermore, for the DNA sample 7739 WGS was also performed at 60× on a most recent HiSeq X Ten sequencer (Illumina) by V4 using Illumina's TruSeq Nano DNA Sample Preparation Kit, which was not polymerase chain reaction (PCR)-free but at the time of library preparation the only kit compatible with the HiSeq X Ten system, according to the manufacturer's instructions for 350-bp insert size.

### Data analysis

Raw data processing, sequence read alignment from FASTQ to BAM format and variant calling to generate VCF files were performed by the four vendors (Supplementary Table S3). For our downstream analyses, aligned BAM files with removed duplicated reads were used, which were either directly provided by vendors (V3 and V4) or deduplicated by us (V1 and V2) using Picard tools version 1.108 or 1.118 (http://picard.sourceforge.net). To determine the

**Table 1.** Experimental design and characteristics of DNA samples used in this study

| Sample # | Gender | DNA source | Extraction method | Purified | WES/Vendor[a] | WGS/Vendor (coverage) |
|---|---|---|---|---|---|---|
| 44 | female | blood | Qiagen column | no | | no WGS |
| 280 | female | blood | Chemagen | no | | no WGS |
| 326 | female | fibroblasts | Chemagen | no | Agilent/V1, V2 | no WGS |
| 2905 | male | blood | Chemagen | yes | NimbleGen/V1, V3 | no WGS |
| 7344 | female | blood | Chemagen | yes | Illumina/V1, V4 | HiSeq/V3 (30×) |
| 7739 | female | saliva | Chemagen | yes | | HiSeq/V1 (60×), V3 (30×), V4 (30×) and XTen/V4 (60×) |

Qiagen column, DNA extraction using Qiagen QIAamp DNA Mini Kit; Chemagen, DNA extraction using PerkinElmer Chemagic Magnetic Separation Module I; Purified, purification of the extracted DNA by re-extraction using Qiagen QIAamp DNA Mini Kit; WES Agilent, SureSelect Human All Exon kit v5+UTR; WES NimbleGen, SeqCap EZ Exome (v3) +UTR; WES Illumina, Nextera Rapid Capture Expanded Exome; V1–V4, vendors 1–4; WGS HiSeq, TruSeq DNA PCR-Free Sample Preparation Kit on a HiSeq2000/2500 system; WGS XTen, TruSeq Nano DNA Sample Preparation Kit on a HiSeq X Ten system.
[a]For all six samples.

number of reads and the coverage at defined minimal read depth (1, 10, 15 and 20×) of platform-specific target regions, RefSeq exons and the subset of exons analysed by Sanger sequencing and the ACE platform of Personalis, we used the SeqMonk program version v0.25.0/v0.26.0/v0.27.0 (http://www.bioinformatics.babraham.ac.uk/projects/seqmonk) with deduplicated BAM files and the settings 'single-end reads' for data import to enable the counting of individual reads and 'remove exact duplicates' for the option 'feature probe generator' (Supplementary Tables S4–S15). Genomic positions of target regions were downloaded from the platforms' websites, whereas genomic coordinates of RefSeq exons were obtained from the UCSC Genome Browser (http://genome.ucsc.edu/, version December 2013). The GC content of RefSeq exons was calculated using the Galaxy platform (http://usegalaxy.org). Data from the X and Y chromosomes were included in our data analyses. For arithmetic means, upper and lower confidence limits (95% confidence intervals) were calculated using critical values of paired *t*-test distribution ($P = 0.05$) and indicated where appropriate.

In order to ensure the comparability of the platforms' mutant (non-reference) allele enrichment performance, we restricted our analyses of the provided VCF files (filtered, if available, Supplementary Table S3) to shared sequence variants targeted by the design of each platform and located in RefSeq coding exons completely covered at 20× by all six platform-vendor combinations. The restriction to such shared sequence variants should largely exclude possible false-positive allele calls. Moreover, in order to avoid the influence of the vendors' different data analysis workflows, we generated genome VCF (gVCF) files, which store sequencing information for both variant and non-variant positions, by applying the same in-house bioinformatics pipeline to all six FASTQ files provided by V1–V4 for each DNA sample (Supplementary Table S3). These gVCF files were filtered to include only positions with ≥20 reads and >30 quality scores in all samples and a non-reference allele (relative allele proportion 10–90%) detected in at least one of the six platform-vendor combinations within the platform's target region and 50-bp flanking sequences.

From our Sanger sequencing data, a total of 78 different single nucleotide variants (SNVs) and 11 different indels are known to be heterozygous in at least one of the six DNA samples, including clinically relevant, disease-causing mutations (Supplementary Tables S1 and S16). Thirty-five variants (30 SNVs and 5 indels) are located within our region of interest (ROI) for clinical sequencing, which consists of the entire protein-coding exonic region with −50 and +20-bp flanking intronic sequences (Supplementary Figure S4) and five SNVs affect UTR (Supplementary Table S1). Using deduplicated BAM and unfiltered VCF files (Supplementary Table S3), these variant positions were analysed for read depth, calling and fraction of non-reference (alternative, mutant) allele and GC content of 30-bp flanking sequences (Supplementary Figures S5–S11 and Supplementary Tables S17–S20).

In addition, three samples (44, 7344 and 7739) were run on a NimbleGen CGH/LOH 6 × 630K array (Roche) according to the manufacturer's instruction. This array contains probes for a total of 501 common SNVs within RefSeq coding exons. Array data of these SNVs located within the designed target region of each platform and exons completely covered at 20× by all six platform-vendor combinations (93, 101 and 53 SNVs for samples 44, 7344 and 7739, respectively) were compared to the corresponding WES variant calls in the provided unfiltered and recalibrated VCF files (Supplementary Table S3). In this comparison, array positions with no array results or false-negative or false-positive calls in all three platforms were excluded (81, 93 and 39 array positions remained for the samples 44, 7344 and 7739, respectively).

The copy number variant (CNV) detection capability of the platforms was assessed by comparing the relative read depth of exons with known deletions (i.e. one copy) to the read depth of normal flanking exons (i.e. two copies) in affected samples and controls using Integrative Genomics Viewer (IGV, Broad Institute). In addition, we also assessed the detection of these deletions by applying the WES-specific CNV calling tools cnMOPS (21) and XHMM (22) for appropriate samples as well as the WGS-specific algorithm BreakDancer (23) for corresponding chromosomes in all four PCR-free WGS datasets. All three CNV detection tools were used with default settings according to the developers' instructions. Moreover, we calculated normalized relative base counts for RefSeq exons on autosomes according to MLPA data analysis (19). In details, the base counts of 21 769 exons completely (100%) covered at 20× in all 36 pro-

vided WES BAM files (i.e. in all combinations of the three platforms, six DNA samples and two vendors) were used for normalization. Copy number calculations were performed for the samples 44, 280, 2905, 7344 and 7739 relative to the sample 326, thereby only considering exons that in sample 326 achieved a coverage of 20× for at least one base and a total base count of ≥1000 in order to reduce the misleading effect of incorrectly mapped reads. Using this calculation, in order to further evaluate the CNV detection properties of the three updated exome enrichment platforms, the relative base counts of 182 exons in 12 different genomic regions with copy numbers known from a NimbleGen CGH 2.1M/4.2M array (Roche, Supplementary Table S21) were assessed. In WGS, using the same 21 769 exons for normalization as in WES, the reproducibility of copy number calculation was assessed based on the base counts of sample 7739 and five additional DNA samples sequenced at 60× on a HiSeq X Ten system as performed for sample 7739.

## RESULTS

### Platform design

In contrast to the designs (target regions) of previous platform versions, NimbleGen now offers the largest target region including coding and untranslated regions (UTR), covering 96 Mb (64 Mb coding + 32 Mb UTR) compared to 75 Mb (50 Mb coding + 25 Mb UTR) of Agilent and 62 Mb (42 Mb coding + 20 Mb UTR) of Illumina. Moreover, NimbleGen promises to capture the largest portion of the entire RefSeq coding exome (98%) followed by Illumina with 95% and Agilent with <94% (Table 2, Supplementary Figure S12 and Supplementary Table S2). Almost each individual RefSeq exon is completely targeted by the designs of NimbleGen (98%) and Illumina (94%) but, notably, in nearly half of the cases (46%) only partially covered by the design of Agilent (Supplementary Table S2).

### Exome enrichment performance

Platform designs may let users expect certain WES performances. However, the question is whether expectations derived from platform designs meet the real laboratory performances. In WES, captured DNA fragments sequenced with sufficient quality result in reads which can be aligned to the reference genome sequence (i.e. mapped), producing an appropriate alignment (BAM) file per DNA sample. In this study, for the six platform-vendor combinations, on average 55.2–90.4% of the raw reads remained after mapping and deduplication, with duplicates accounting for 8.4–37% of the mapped reads. Applying the same platform, the number of deduplicated mapped reads was significantly different between V1 and V2 (Agilent) as well as between V1 and V4 (Illumina). This difference is at least partially due to variation in the proportion of duplicated reads, suggesting the impact of laboratory workflow on WES (Figure 1A, Supplementary Figure S13 and Supplementary Tables S22–S24). Indeed, duplicated reads may rather depend on a laboratory workflow, whereas variation in the proportion of unaligned reads may rather be due to alignment tools as demonstrated by comparing provided and in-house generated BAM files (Supplementary Figure S13). Off-target enrichment was assessed as the fraction of total aligned reads which mapped more than 500 bp outside the designed target regions. As in 2011 (3), Illumina showed the highest proportion of off-target reads (~40%) compared to Agilent and NimbleGen (Figure 1A and Supplementary Tables S4 and S5).

While previous versions of NimbleGen showed the highest enrichment efficiency (2–5), the current version of NimbleGen was only able to achieve best exome coverage in some samples and vendor (Figures 1 and 2A). Similarly, also Illumina showed distinct inter-sample and inter-vendor variation. In contrast, for the Agilent platform the high enrichment efficiency of target and coding regions was stable between the two respective vendors and among all six DNA samples (Figures 1 and 2, Supplementary Figure S14). The latter result we confirmed by analysing 24 additional DNA samples extracted from blood or saliva (Supplementary Figure S3). This high enrichment performance and superior robustness of Agilent represents a clear improvement of previous versions (Table 2 and Supplementary Figures S1 and S2).

The comparison of the enrichment performances of the three platforms used with the same experimental setting by the same vendor (V1) revealed that Agilent reached the expected mean read depth of 100 (100×) for RefSeq coding region (105 ± 17), 94.7 ± 1.2% of which were sequenced at 20×, i.e. with a read depth of at least 20 (Figure 1B and Supplementary Table S6). NimbleGen with a mean read depth of 93 ± 17 showed the lowest 20× coverage of RefSeq coding region (79.0 ± 3.5%), whereas Illumina with lowest mean read depth (79 ± 11) overall achieved significantly higher 20× RefSeq coverage (87.0 ± 2.3%). Similar results for read depth and overall coverage performance was revealed by V1 also for the platforms' target regions (Figure 1A and Supplementary Table S4). The discrepancy between read depth and coverage performance observed for NimbleGen and Illumina is most likely due to unequal read distribution, which can be driven by the GC content of genomic sequences (24).

The uniformity of the coverage of RefSeq coding exons was assessed by calculating the fraction of exons reaching an average read depth within ±70% of mean read depth over all coding exons (2). With the highest mean read depth, Agilent reached a fraction of ~80% regardless of vendor. Similar results were obtained for Illumina (~75%). In contrast, NimbleGen showed a high inter-vendor variation with ~62% for V1 and ~93% for V3 (Figure 1B and Supplementary Table S8), whereas WGS resulted in superior uniformity of the coverage of RefSeq coding exons with at least 97% (Supplementary Table S9).

In addition, we assessed what proportion of RefSeq coding exons are completely (i.e. 100%) covered at ≥20× and hence suitable for clinical WES (5). Regardless of vendor, Agilent performed best, however still far from 100%, with an average of 86.7 ± 3.3% (V1) and 92.8 ± 1.0% (V2) followed by NimbleGen with 72.4 ± 3.4% (V1) and 85.9 ± 14.1% (V3) as well as by Illumina with 63.1 ± 6.1% (V1) and 53.6 ± 4.2% (V4). Only 28.7 ± 5.8% of all RefSeq coding exons are completely covered at ≥20× by all three platforms (Figure 3A and Supplementary Tables S8 and S10). In fact, Agilent exceeded the expectation of complete (100%) exon

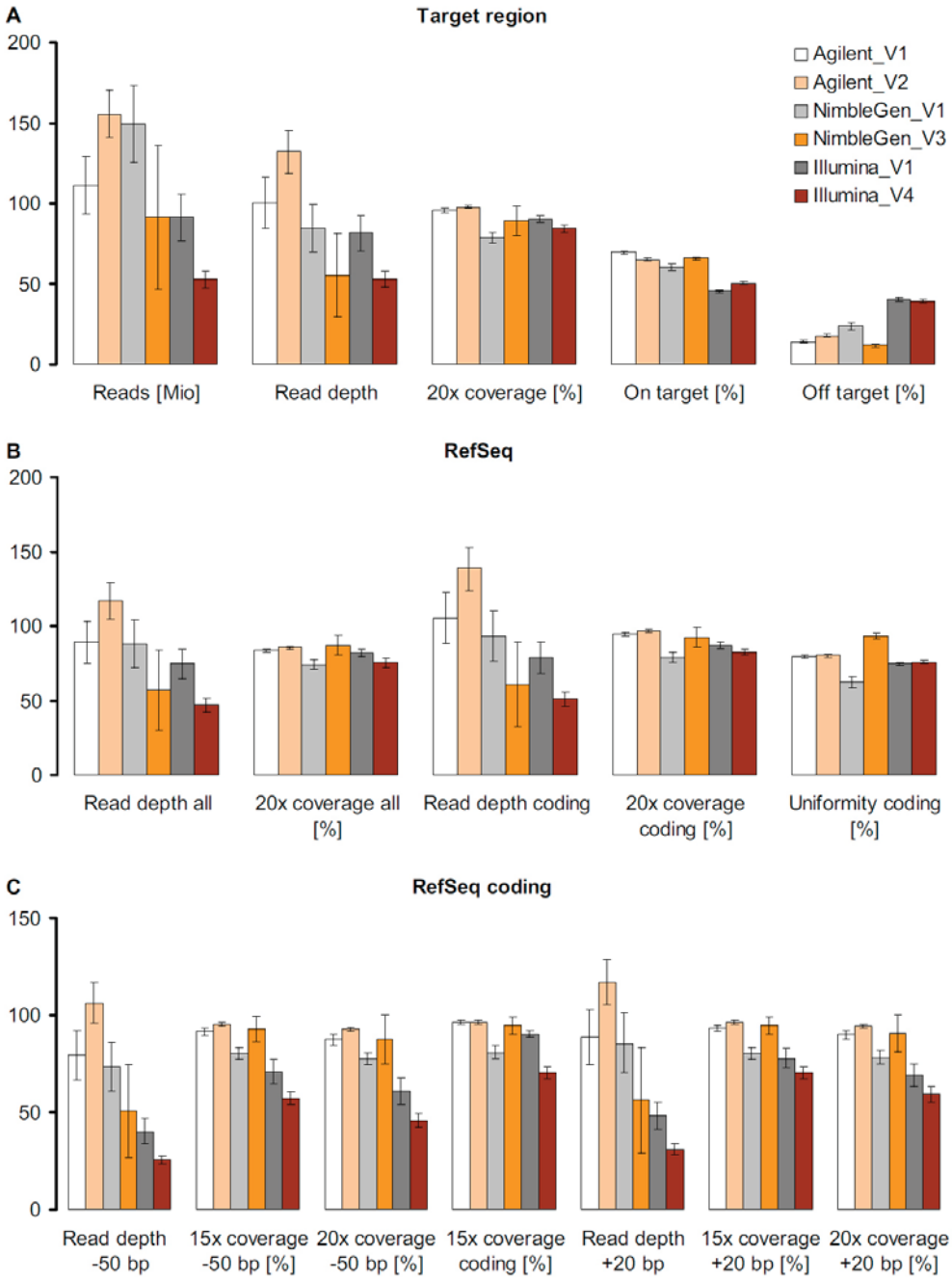



**Figure 1.** Enrichment efficiency of the three updated exome enrichment platforms (Agilent, NimbleGen and Illumina) performed by four vendors (V1, V2, V3 and V4). (**A**) Mean number of aligned reads (as million reads), mean read depth and percentage of coverage at 20× for each designed target region as well as mean percentage of on-target reads (i.e. within designed target regions) and mean percentage of off-target reads (i.e. within regions more than ±500 bp outside the designed target regions). Note that values for aligned reads indicate the total number of mapped reads without duplicates for V1 and V2 and only uniquely mapped reads without duplicates for V3 and V4 (Supplementary Table S3). (**B**) Mean read depth and percentage of coverage at 20× for all and only coding exons of the RefSeq database as well as uniformity of the coverage of RefSeq coding exons calculated as the fraction of exons reaching an average read depth within ±70% of mean read depth over all coding exons (uniformity coding). (**C**) Mean read depth and percentage of coverage at 15 and 20× for RefSeq coding exons as well as for −50-bp and +20-bp flanking intronic regions. Given are means of all six DNA samples ($n = 6$); error bars indicate 95% confidence intervals. Values were calculated using the SeqMonk program (http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/) and are presented in Supplementary Tables S4-S8 and S12–S13. For complete coverage of RefSeq coding exons see Figure 3.

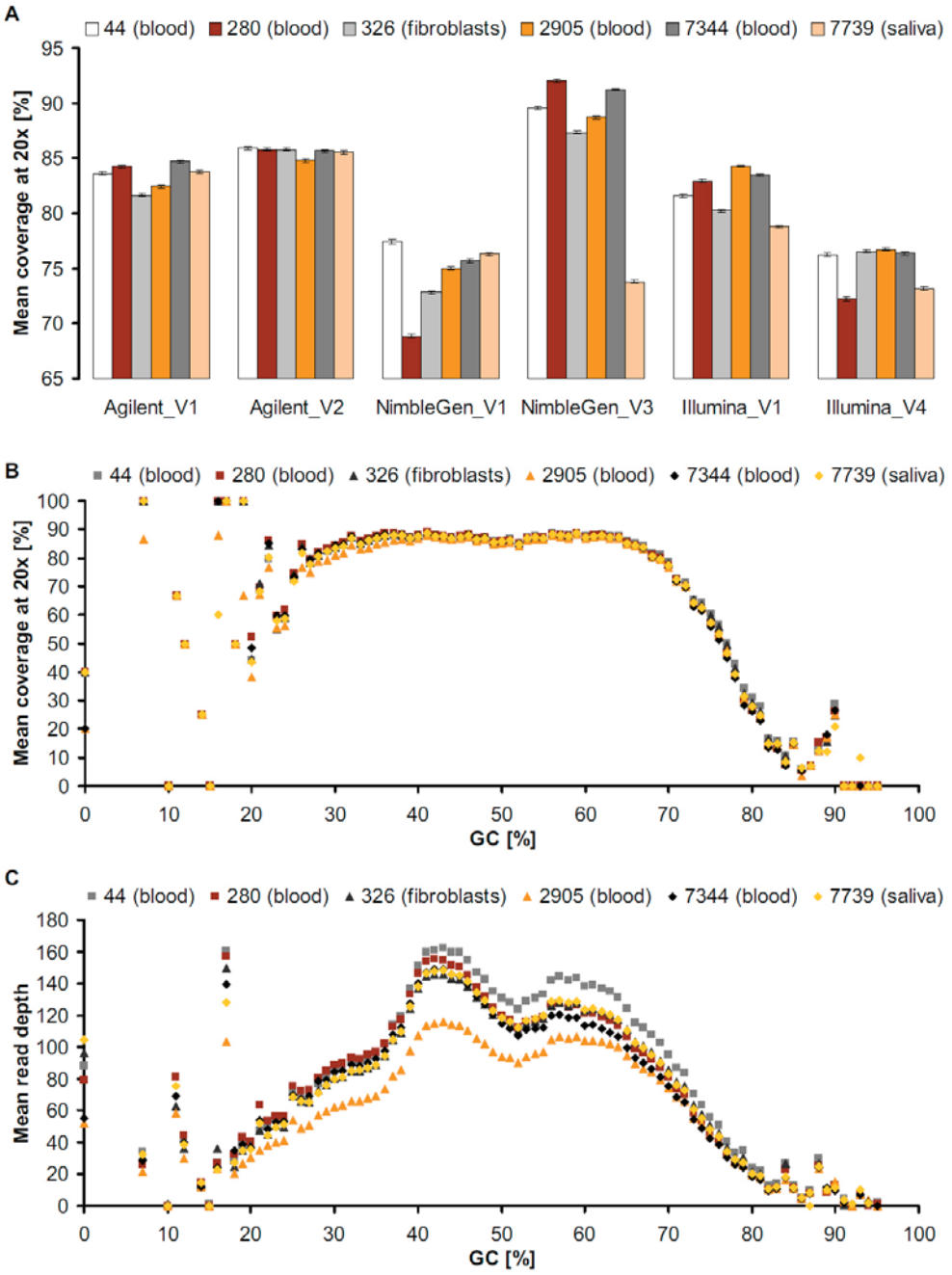



**Figure 2.** Differences among DNA samples. (**A**) Mean coverage of RefSeq exons ($n = 233\,644$) at 20× (expressed in percentage of the entire exon length) for all six platform-vendor combinations and DNA samples (44, 280, 326, 2905, 7344 and 7739) derived from blood, fibroblasts or saliva. Values were obtained by using the SeqMonk program (www.bioinformatics.babraham.ac.uk/projects/seqmonk) and are presented in Supplementary Tables S6 and S12. Error bars indicate 95% confidence intervals for the arithmetic means of all corresponding exons. (**B** and **C**) Mean coverage at ≥20 reads (B) and mean read depth (C) of RefSeq exons per GC content for each DNA sample exemplified by the WES data of V2 using Agilent, demonstrating its high performance stability across samples.

**Table 2.** Overview of studies evaluating exome enrichment platforms as well as summary of which of the platforms performed best for the assessed aspects

| | This study | Clark *et al.* 2011 (3) | Asan *et al.* 2011 (2) | Parla *et al.* 2011 (4) | Sulonen *et al.* 2011 (5) | Chilamakuri *et al.* 2014 (6) |
|---|---|---|---|---|---|---|
| Enrichment platforms | Agilent v5+UTR, NimbleGen v3+UTR and Illumina Nextera Expanded Exome | Agilent v3, NimbleGen v2 and Illumina TruSeq Exome | Agilent v1, NimbleGen v1 (in-solution), 2.1M array | Agilent v1 and NimbleGen v1 | Agilent v1, v3 and NimbleGen v1, v2 | Agilent v4, NimbleGen v3, Illumina TruSeq Exome and Illumina Nextera Expanded Exome |
| Sequencing platform | Illumina HiSeq 2000/2500 paired-end 100-bp reads | Illumina HiSeq 2000 paired-end 100-bp reads | Illumina HiSeq 2000 paired-end 90-bp reads | Illumina GAIIx, paired-end 76-bp reads | Illumina GAIIx, paired-end 82-bp reads | Illumina HiSeq 2000 paired-end 100-bp reads |
| DNA samples | Six samples performed by different vendors, 24 samples performed by one vendor using Agilent | One sample | One sample | Six HapMap samples (two for both platforms and four only for NimbleGen) | One sample for all platforms, 25 samples for one platform | One sample with two technical replicates per platform |
| Region for sequence variant calling | Common designed target region in RefSeq coding exons 100% covered at 20× by all platform-vendor combinations | Genome-wide | Designed target region with 200-bp flanking sequences | CCDS | Genome-wide, designed target region (individual and common), and CCDS | Designed target region (individual and common), CCDS, RefSeq (coding and UTR) and Ensembl |
| Largest designed target region | NimbleGen | Illumina | Agilent | Agilent | Agilent v2 | NimbleGen |
| Largest coding region (reference) | NimbleGen (RefSeq) | Agilent (RefSeq, Ensembl CDS) | Agilent (CCDS) | Agilent (CCDS) | Agilent v2 (CCDS) | Illumina (CCDS, RefSeq, Ensembl) |
| Best designed target enrichment efficiency | Agilent | NimbleGen | NimbleGen (array and in-solution) | NimbleGen | NimbleGen v2 | Agilent |
| Lowest off-target enrichment | Agilent and NimbleGen | NimbleGen | NimbleGen (array and in-solution) | NimbleGen | NimbleGen v1 | Agilent and NimbleGen[a] |
| Best GC-rich region enrichment | Agilent | Agilent | NimbleGen array | No data | NimbleGen v2 | Illumina Nextera |
| Highest accuracy of SNV detection (benchmark) | Agilent (Sanger sequencing, MLPA and SNP array) | Agilent (SNP array) | No clear difference among platforms (SNP array and WGS) | Agilent (HapMap and 1000 Genome Project data) | NimbleGen v2 (SNP array) | No determination of accuracy by comparison to a benchmark (only calling of SNVs) |

[a]Estimated from provided figures, as off-target reads were reported as relative proportion of filtered reads rather than total mapped reads; CCDS, Consensus Coding Sequences.

coverage from the capture design (cf. only 54.2% of the RefSeq exons are covered 100% by the target design of Agilent), whereas NimbleGen (98.3% expected) and Illumina (94.0% expected) failed to reach their promised coverage (Figure 3A and Supplementary Table S2). Our data suggest that the proportion of incompletely covered exons can be reduced by combining platform, rather than vendor, performances. Indeed, the combination of the two best (Agilent and NimbleGen) and all three platforms left only $2.3 \pm 0.5\%$ and $1.7 \pm 0.3\%$ of the RefSeq coding exons uncovered at $\geq 20\times$, respectively, whereas the combined performance of both vendors using Agilent (V1 and V2) could only reduce the proportion of not completely covered exons from $7.2 \pm 1.1\%$ (Agilent alone by V2) to $6.8 \pm 0.9\%$ (Figure 3B and Supplementary Table S10). Alternatively, particularly focused enrichment can help to improve the standard WES coverage of exons of interest as exemplified using a clinical exome platform (Supplementary Table S11). However, in comparison to all exome enrichment platforms used in this study, WGS ($60\times$) showed fewer uncovered exons at comparable read depth (i.e. at 10–15×; Supplementary Tables S9 and S11).

To assess the impact of GC content on enrichment performance, we plotted mean read depth as well as mean coverage at $20\times$ against the GC content of RefSeq exons (Figures 2B, C and 4 and Supplementary Figures S15 and S16). Both Agilent and Illumina suggested correlation between GC content and read depth regardless of vendor, whereas NimbleGen resulted in clear differences between vendors and when performed by V3, also among samples. Read depth of exons with very low (<20%) or high (>80%) GC content was variable or low for all platforms, however, Agilent thereby performed more robust and slightly better. This limitation of WES in capturing GC-rich regions also differently affected previous platform versions (Table 2) and can be overcome by WGS, as it is free from genomic hybridization/capture, especially by PCR-free WGS. Indeed, our PCR-free WGS data showed no distinct negative effect of high or low GC content on the mean read depth of RefSeq exons, covering GC-rich regions (>80%) much better than WES. Non-PCR-free WGS (HiSeq X Ten) achieved slightly lower enrichment of GC-rich exons but the observed bias was far less pronounced than in WES (Figure 5 and Supplementary Figure S17).

### Enrichment and detection of non-reference alleles

Sequence variant detection by WES requires both the experimental enrichment and bioinformatics calling of mutant alleles. As read alignment and variant calling can considerably influence sequence variant detection (http://genomeinabottle.org; http://www.bioplanet.com/gcat; 25), we focussed on data generated by the same bioinformatics workflow for the assessment of the enrichment of non-reference (mutant) alleles. Accordingly, we focused on filtered VCF files generated by the bioinformatics workflow of V1 for all three platforms as well as gVCF files created by our in-house data analysis pipeline for all six platform-vendor combinations. We restricted the analysis of the provided VCF files to RefSeq coding regions, whereas we created our gVCF files for the entire target region of each platform including 50-bp flanking sequences. In both cases, for each DNA sample only non-reference alleles covered and called by all six platform-vendor combinations were considered in order to reduce the possibility of miscalling. In the provided VCF files, by assessing the mean relative propor-

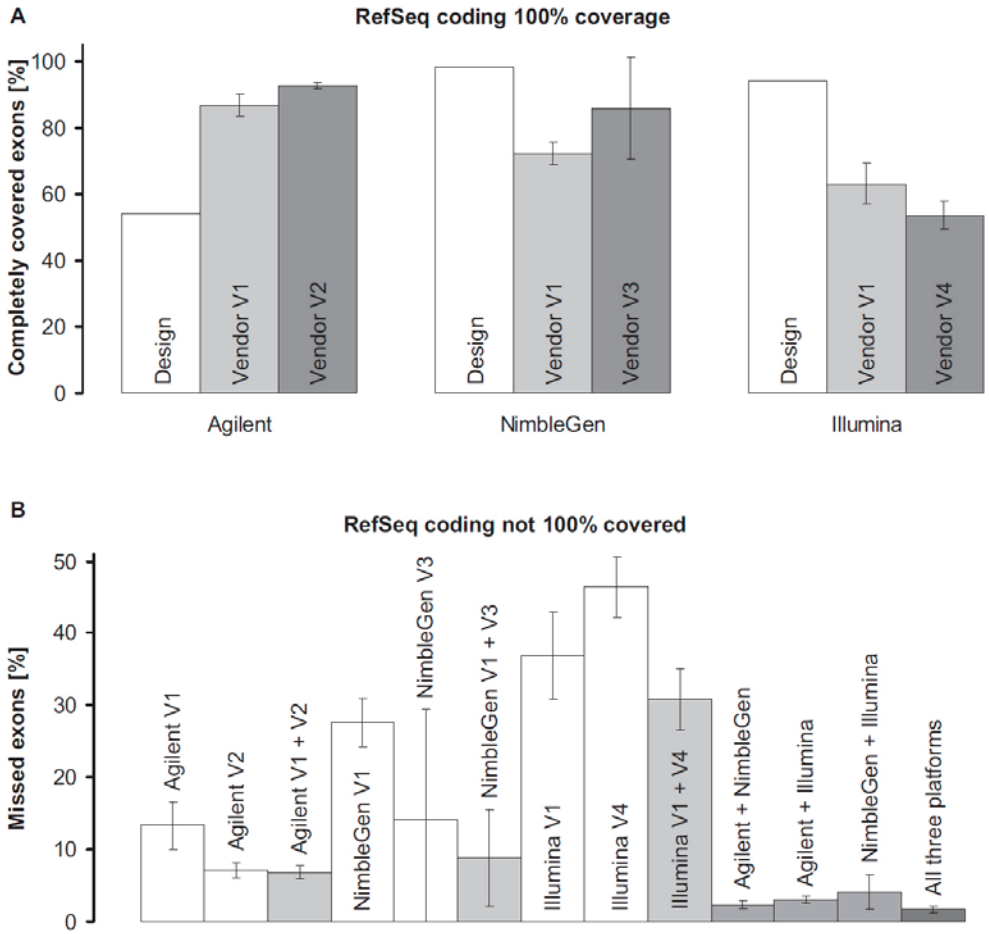



**Figure 3.** Complete (i.e. 100%) coverage of RefSeq coding exons. (**A**) Proportion of RefSeq coding exons 100% covered by each designed target region (design) and by ≥20 reads effectively produced by each vendor (vendors V1–V4). (**B**) Proportions of RefSeq coding exons not 100% covered at 20× (missed exons). If not otherwise indicated, data of all corresponding vendors are included. Given are means of all six DNA samples ($n = 6$); error bars indicate 95% confidence intervals. Values were calculated using the SeqMonk program (http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/) and are presented in Supplementary Tables S2, S8 and S10.

tion of non-reference alleles (26), no considerable difference was observed among platforms neither for all variants nor for indels only, suggesting comparable sensitivity to the detection of mosaicism (Figure 6). Nevertheless, the enrichment of non-reference alleles was more stable for Agilent, resulting in reproducibly lower variation in the relative proportions of alternative alleles compared to NimbleGen and Illumina (Figure 6, Supplementary Figure S18 and Supplementary Table S25). Moreover, as one might expect considering hybridization mismatches, the enrichment of non-reference alleles was rather lower (<50%) than the capture of reference ones (>50%). All these results derived from the analysis of provided VCF files are supported by comparable findings obtained from heterozygous SNVs characterized by Sanger sequencing (Supplementary Figure S10) as well as from shared heterozygous SNVs and indels in our in-house generated gVCF files (Supplementary Figures S19 and S20 and Supplementary Table S26).

To determine the accuracy of WES variant detection, we analysed heterozygous SNVs and small indels, including clinically relevant mutations, previously characterized by Sanger sequencing in our DNA samples. Unlike previous studies (Table 2), we restricted this analysis to variants in a ROI relevant for clinical sequencing, which comprises exons with −50-bp and +20-bp flanking intronic sequences, as well as to UTR variants. This Sanger-based benchmarking provided not only the most accurate reference genotypes (25) but also allowed us to compare WES with Sanger sequencing with respect to clinical use in our set of genes. For all assessed heterozygous exonic positions, a minimum read depth of 20 was achieved by all three updated platforms. The coverage of intronic positions, however, varied among platforms and DNA sources similar to the perfor-

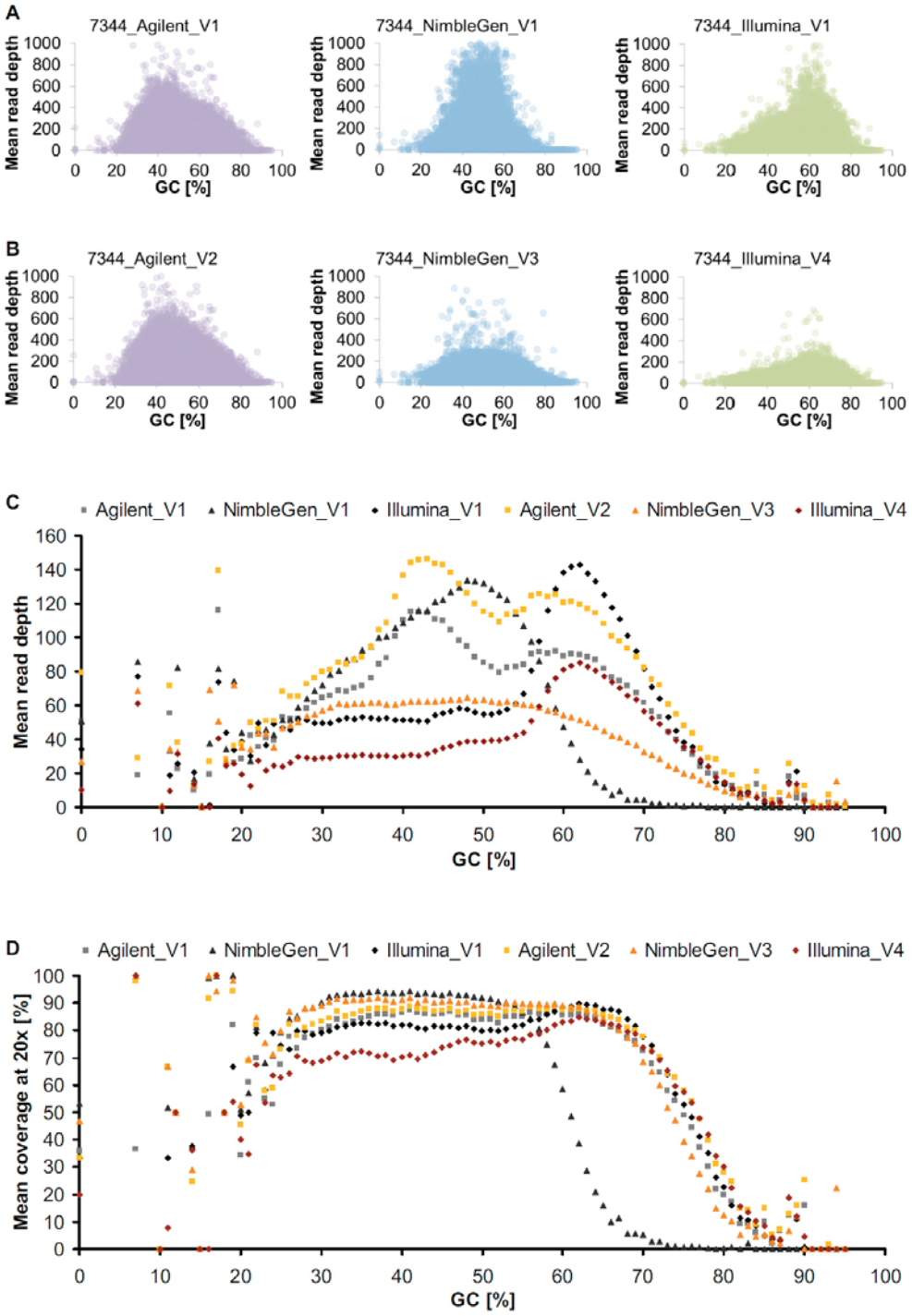



**Figure 4.** Differences in sensitivity to GC content among all platform-vendor combinations (average of all six DNA samples). (**A** and **B**) Scatter plot showing GC content and achieved read depth of RefSeq exons (coding and UTR) for the three updated exome enrichment platforms performed by the same vendor (V1, A) and different vendors (V2–V4, B), exemplified for sample 7344 (plots of all six samples are shown in Supplementary Figures S15 and S16). (**C**) Mean read depth of RefSeq exons per GC content shown as means of all samples. (**D**) Mean 20× coverage of RefSeq exons per GC content shown as means of all samples.

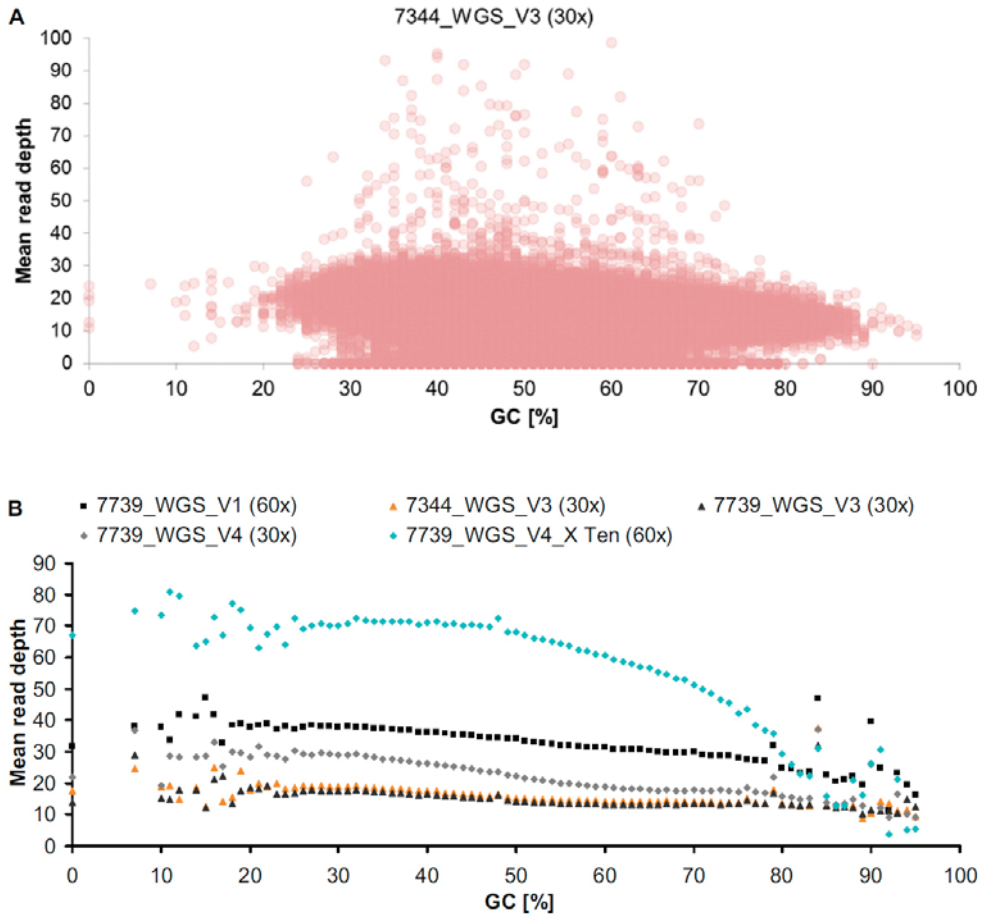



**Figure 5.** Influence of GC content on mean read depth in WGS. (**A**) GC content and achieved read depth of RefSeq exons (coding and UTR) exemplified by WGS of sample 7344 performed by V3 (plots of all WGS datasets are shown in Supplementary Figure S17). (**B**) Means of read depths of RefSeq exons per GC content. X Ten, HiSeq X Ten system.

mance observed genome-wide (Supplementary Figures S5 and S6). Regardless of vendor and hence also of mapping and variant calling workflow, Agilent correctly detected all our SNVs and called no false-positive variants, whereas NimbleGen (V1) failed to detect one SNV in UTR and Illumina (V4) identified three false-positive heterozygous SNVs (Supplementary Tables S17–S19). By assessing calling accuracy for indels, which was not analysed in previous studies (2–6), we found that all three platforms failed to detect some small indels, although Agilent identified the highest number of indels in ROI (Supplementary Table S20). All analysed SNVs covered by array data (81, 93 and 39 for samples 44, 7344 and 7739, respectively) were called correctly in WES regardless of platform and vendor (Supplementary Table S27).

Moreover, our gVCF files generated by using the same bioinformatics pipeline also allowed a comparative assessment of variant detection in all six platforms-vendor combinations. We thereby focussed on positions with ≥20 reads and >30 quality scores which were called as heterozygous by only one or by all but one platform and hence indicate possibly false-positive or false-negative variant calls. Agilent resulted in the lowest proportion of such putative calling errors (0.82 ± 0.04% of all >14 000 analysed variant positions) followed by NimbleGen (1.52 ± 0.10%) and Illumina (1.66 ± 0.19%) (Supplementary Figure S21 and Supplementary Table S28). However, high quality variants (≥20 reads and quality >30) called by all platforms may not be sufficient to exclude all putative alignment errors and other variant calling pipelines may also lead to different results. Indeed, WES users should be aware that for the same raw reads (FASTQ file) different read aligner and variant caller combinations may result in considerably different number of true and false variants (Genome Comparison & Analysis Testing at http://www.bioplanet.com/gcat). The assessment of the effect of bioinformatics pipelines on WES variant detection, however, is beyond the scope of this study and should be addressed by further works focusing on this topic.

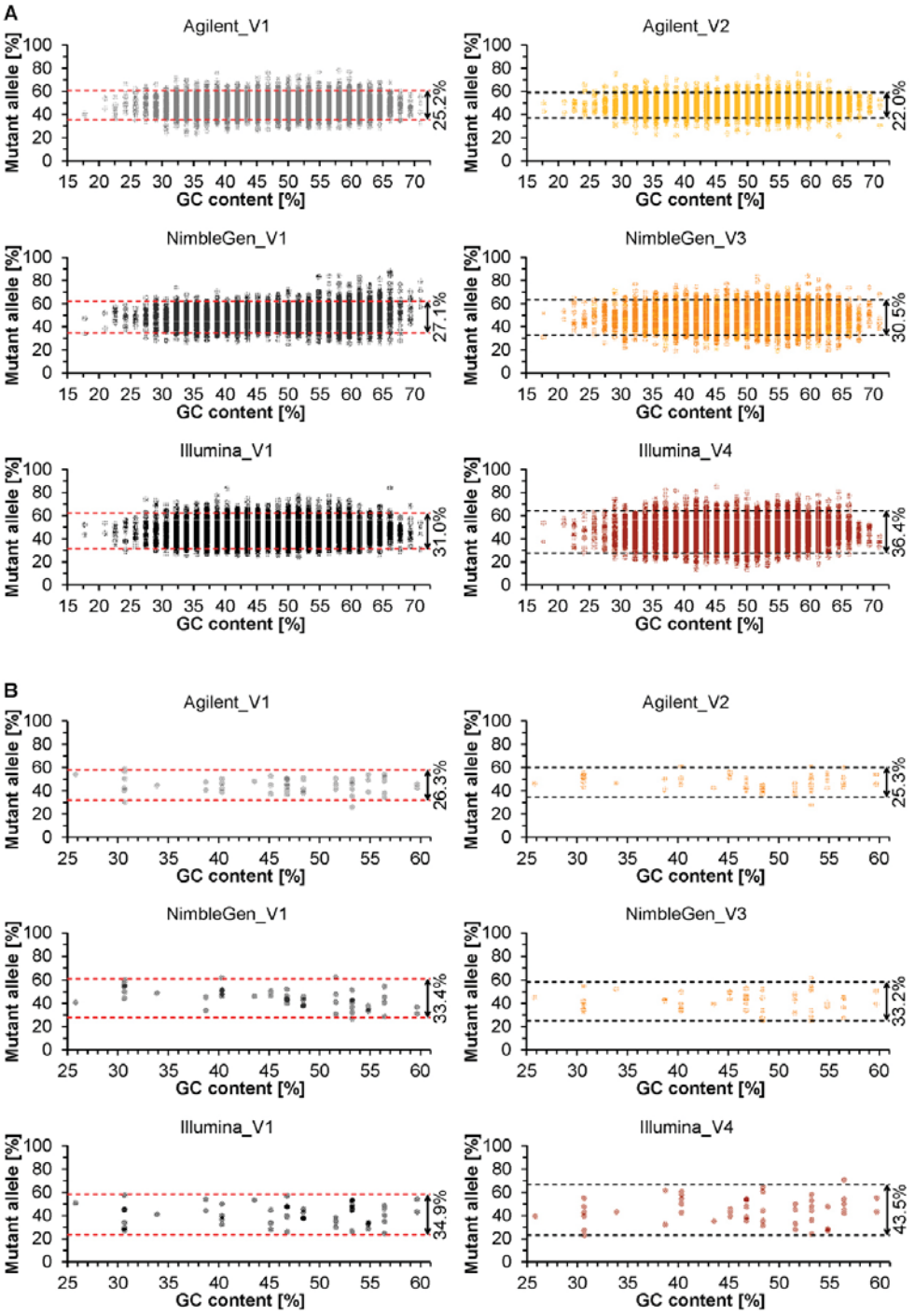



**Figure 6.** Relative proportions of non-reference (mutant) alleles called in the VCF files provided by vendors (V1–V4). The analysis was restricted to shared heterozygous variants within the designed target regions of the three platforms (Agilent, NimbleGen and Illumina) located in exons completely (100%) covered at 20× by all six platform-vendor combinations. (**A** and **B**) Heterozygous SNVs (A) and indels (B) listed according to GC content of 30-bp flanking sequences (for indel lengths see Supplementary Figure S18). Shown are values of all six DNA samples. Dashed lines indicate an interval within which 95% of the relative proportions of non-reference alleles lie (calculated according to the Student's *t* distribution as the mean of *n* percentage values ± critical *t*-value ($t_{crit,n-1}$) × *SD* using $n = 8\,687$, $t_{crit} = 1.960$ and $n = 51$, $t_{crit} = 2.009$ for A and B, respectively).

The quantitative, CNV detection properties of the platforms (27,28) were not evaluated in previous studies (2–6). Thus, by using IGV we analysed one DNA sample (sample 44) that harbours a previously characterized heterozygous 27-kb deletion affecting the complete exon 1 of the *FBN1* gene (19). Compared to the other five samples captured by using Agilent and Illumina, the read depth of the deleted exon 1 was distinctly lower than the read depth of the flanking undeleted exons 2–5. For NimbleGen, however, the distribution of read depth among samples and exons was less stable (regardless of alignment tool as confirmed by our in-house generated BAM files) and thus the heterozygous one-exon deletion was not clearly detectable (Supplementary Figure S22). Moreover, by using IGV we analysed WES and WGS data for a 7-kb deletion to compare their potential to detect large deletions (Supplementary Figure S23). Both sequencing strategies allowed the detection of large deletions by a reduction of read depth compared to flanking exons and sequences, respectively. However, by using WGS a more accurate determination of breakpoint positions is possible, whereas by using WES additional methods may be needed (29). Using CNV calling tools, none of the two tested deletions were detected in the WES datasets by the WES-specific cnMOPS (21) and XHMM (22) methods. In contrast, the 7-kb deletion was called in all three PCR-free WGS data sets (V1, V3 and V4) using the algorithm BreakDancer (23) (data not shown).

At genome-wide level, we extended the evaluation of the CNV detection properties of the three updated exome enrichment platforms not only for a large number of exons and different DNA samples but also for more known CNVs. The comparison of the relative base counts of 21 769 RefSeq exons completely covered at 20× in all 36 platform-vendor-sample combinations showed the lowest variation for Agilent regardless of vendor and revealed comparable variation between WES and non-PCR-free WGS (Supplementary Figures S24 and S25 and Supplementary Table S29). The lowest variation of Agilent and thus its potential for best CNV detection was confirmed by assessing 182 exons with copy numbers known from array CGH (Supplementary Figure S26 and Supplementary Table S30).

## DISCUSSION

Users of WES expect coverage of the entire coding region of all known genes and sufficient read depth for the covered regions. This comparative study provides the most recent and comprehensive data to answer the question of which current standard WES enrichment platform is most suitable to meet these expectations. By including different sequencing providers (vendors) and samples into our performance comparison, our study design allowed us to evaluate the three most recent standard exome enrichment platforms not only within the same experimental and bioinformatics setting but also between different settings and among different DNA sources. Our study focuses on the enrichment of both reference and non-reference alleles, keeping the influence of bioinformatics workflow, which is a matter of ongoing research, as low as possible. We observed that the Agilent SureSelect Human All Exon v5+UTR enrichment platform is superior to the other two platforms with regard to overall performance and robustness. This superior performance of the most recent Agilent platform is, however, not applicable to its previous versions as shown by comparisons in 2011 (Table 2) and our preliminary study (Supplementary Figures S1 and S2).

Although the NimbleGen platform is designed to enrich the largest target region and the highest proportion of the coding region of the genome, the most recent version of the Agilent platform produces higher and more consistent exome coverage. This discrepancy may at least partially be explained by different exome designs. Whereas the designed target regions of NimbleGen and Illumina represent the region intended to be enriched, the target region specified by Agilent only includes sequences effectively covered by hybridization probes. Thus, the designed exome coverage does not completely reflect effective hybridization probe coverage and exome enrichment efficiency (Figure 3A, Supplementary Figures S4 and S12, and Supplementary Table S2). This illustrates that information on the designed target region provided by platform selling companies may be misleading for WES users, emphasizing the need for laboratory evaluation of real platform enrichment performances. From a technical point of view, the performance of Agilent may also, at least partially, be explained by the relatively long RNA baits of this platform. It appears that longer RNA baits lead to better hybridization and enrichment efficiency as well as tolerate larger hybridization mismatches and thus provide more stable post-capture representation of non-reference alleles harbouring sequence variants, especially indels (Figure 6, Supplementary Figures S10 and S18–S20, and Supplementary Table S20). However, we observed that the higher hybridization efficiency and mismatch tolerance of Agilent does not result in increased, unspecific off-target capture (Figure 1A).

Alternatively, NimbleGen may be the enrichment method of choice for users interested in regions exclusively covered by this platform. However, users should be aware of more pronounced bias for GC-rich regions and potentially reduced enrichment efficiency due to sensitivity to vendor and DNA sources (Figures 1, 2 and 4). This bias and sensitivity may be explained, at least partially if not all, by the relatively low hybridization temperature during exome enrichment (47°C compared to 65°C and 58°C in Agilent and Illumina, respectively) and/or by the relatively short size of hybridization probes (55–105 bp DNA compared to 90/120 bp RNA and 95 bp DNA in Agilent and Illumina, respectively; Supplementary Table S2). For applications with limited starting material, Illumina with 50 ng required DNA amount is superior to NimbleGen and Agilent, both of which require more DNA as standard input (Supplementary Table S2). For clinical WES, particularly focused exome enrichment or the combination of Agilent and NimbleGen or of all three platforms can result in 100% sequence coverage of more exons than by using the standard platforms alone. As the three platforms neither alone nor in combination can completely capture all coding exons, the enrichment performance of each platform requires improvement, at least for exons of clinically relevant genes (Figure 3B, Supplementary Figure S27 and Supplementary Tables S9–S11).

Like Sanger sequencing, WES allows the simultaneous analysis not only of coding exons but also of flanking in-

tronic sequences involved in normal splicing as well as the sufficient enrichment of non-reference alleles harbouring SNVs and small indels, all of which are essential for the detection of disease-causing mutations. Thus, the question arises whether or not WES can replace Sanger sequencing in mutation detection (30). Our data suggest that with respect to RefSeq coding exons and flanking intronic sequences none of the three updated WES platforms is suitable to replace Sanger sequencing, although Agilent appears to be more suitable, not least due to its superior robustness (Figure 2). Hence, for best results, particularly in clinical WES, the complete representation and sufficient coverage of each tested gene region has to be ensured, especially for GC-rich regions and deeper intronic positions.

Indeed, WES may fail to capture such regions and have difficulties to detect CNVs. We observed that capture-free, especially PCR-free, WGS can overcome these limitations, making WGS superior to WES, at least in those cases. Furthermore, the exome is not a fixed entity and still subject to changes. Projects such as ENCODE (31) will enhance the interpretation of non-coding, regulatory variants and their importance in genomic research and gene diagnostics will increase. With such changes in knowledge, WES capture design will require constant adaptation and for unsolved cases a re-run of WES will have to be considered, whereas by using WGS the entire genomic information is largely present so that only data analysis has to be repeated. One may assume that WGS will become less expensive, as shown by the recent introduction of the HiSeq X Ten system, and thus be more popular in the near future. However, it remains to be answered whether WGS will complement or replace WES. It is probable that WES with less and better interpretable sequencing data and emerging better enrichment performance will be widely used as an effective alternative to WGS in both research and diagnostics. Nevertheless, the clinical application of these powerful tools should proceed with care and be supported by the patient's health insurance, especially in testing a large number of exons and genes as well as in cases in which no *a priori* knowledge of gene(s) responsible for a particular disease phenotype exists.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank the four sequencing vendors and Personalis Inc. involved in this study for performing exome enrichment, sequencing, mapping and variant calling. We are grateful to the members of the Center for Cardiovascular Genetics and Gene Diagnostics for providing Sanger sequencing data.

## FUNDING

Bangerter-Rhyner-Stiftung; COFRA Foundation; Ebnet-Stiftung; Foundation Suyana; Gebauer Stiftung; Hirzel-Callegari Stiftung; Jubiläumsstiftung Swiss Life. Funding for open access charge: Foundation for People with Rare Diseases.
*Conflict of interest statement.* None declared.

## REFERENCES

1. Ng,S.B., Turner,E.H., Robertson,P.D., Flygare,S.D., Bigham,A.W., Lee,C., Shaffer,T., Wong,M., Bhattacharjee,A., Eichler,E.E. *et al.* (2009) Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*, **461**, 272–276.
2. Asan, Xu,Y., Jiang,H., Tyler-Smith,C., Xue,Y., Jiang,T., Wang,J., Wu,M., Liu,X., Tian,G. *et al.* (2011) Comprehensive comparison of three commercial human whole-exome capture platforms. *Genome Biol.*, **12**, R95.
3. Clark,M.J., Chen,R., Lam,H.Y., Karczewski,K.J., Euskirchen,G., Butte,A.J. and Snyder,M. (2011) Performance comparison of exome DNA sequencing technologies. *Nat. Biotechnol.*, **29**, 908–914.
4. Parla,J.S., Iossifov,I., Grabill,I., Spector,M.S., Kramer,M. and McCombie,W.R. (2011) A comparative analysis of exome capture. *Genome Biol.*, **12**, R97.
5. Sulonen,A.M., Ellonen,P., Almusa,H., Lepisto,M., Eldfors,S., Hannula,S., Miettinen,T., Tyynismaa,H., Salo,P., Heckman,C. *et al.* (2011) Comparison of solution-based exome capture methods for next generation sequencing. *Genome Biol.*, **12**, R94.
6. Chilamakuri,C.S., Lorenz,S., Madoui,M.A., Vodak,D., Sun,J., Hovig,E., Myklebost,O. and Meza-Zepeda,L.A. (2014) Performance comparison of four exome capture systems for deep sequencing. *BMC Genomics*, **15**, 449.
7. Lam,H.Y., Clark,M.J., Chen,R., Natsoulis,G., O'Huallachain,M., Dewey,F.E., Habegger,L., Ashley,E.A., Gerstein,M.B., Butte,A.J. *et al.* (2011) Performance comparison of whole-genome sequencing platforms. *Nat. Biotechnol.*, **30**, 78–82.
8. Gonzaga-Jauregui,C., Lupski,J.R. and Gibbs,R.A. (2012) Human genome sequencing in health and disease. *Annu. Rev. Med.*, **63**, 35–61.
9. Rieber,N., Zapatka,M., Lasitschka,B., Jones,D., Northcott,P., Hutter,B., Jager,N., Kool,M., Taylor,M., Lichter,P. *et al.* (2013) Coverage bias and sensitivity of variant calling for four whole-genome sequencing technologies. *PLoS One*, **8**, e66621.
10. Dewey,F.E., Grove,M.E., Pan,C., Goldstein,B.A., Bernstein,J.A., Chaib,H., Merker,J.D., Goldfeder,R.L., Enns,G.M., David,S.P. *et al.* (2014) Clinical interpretation and implications of whole-genome sequencing. *Jama*, **311**, 1035–1045.
11. Choi,M., Scholl,U.I., Ji,W., Liu,T., Tikhonova,I.R., Zumbo,P., Nayir,A., Bakkaloglu,A., Ozen,S., Sanjad,S. *et al.* (2009) Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 19096–19101.
12. McInerney-Leo,A.M., Marshall,M.S., Gardiner,B., Coucke,P.J., Van Laer,L., Loeys,B.L., Summers,K.M., Symoens,S., West,J.A., West,M.J. *et al.* (2013) Whole exome sequencing is an efficient, sensitive and specific method of mutation detection in osteogenesis imperfecta and Marfan syndrome. *Bonekey Rep.*, **2**, 456.
13. O'Rawe,J., Jiang,T., Sun,G., Wu,Y., Wang,W., Hu,J., Bodily,P., Tian,L., Hakonarson,H., Johnson,W.E. *et al.* (2013) Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing. *Genome Med.*, **5**, 28.
14. Wang,Z., Liu,X., Yang,B.Z. and Gelernter,J. (2013) The role and challenges of exome sequencing in studies of human diseases. *Front. Genet.*, **4**, 160.
15. Yang,Y., Muzny,D.M., Reid,J.G., Bainbridge,M.N., Willis,A., Ward,P.A., Braxton,A., Beuten,J., Xia,F., Niu,Z. *et al.* (2013) Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N. Engl. J. Med.*, **369**, 1502–1511.
16. Rabbani,B., Tekin,M. and Mahdieh,N. (2014) The promise of whole-exome sequencing in medical genetics. *J. Hum. Genet.*, **59**, 5–15.
17. Matyas,G., De Paepe,A., Halliday,D., Boileau,C., Pals,G. and Steinmann,B. (2002) Evaluation and application of denaturing HPLC for mutation detection in Marfan syndrome: Identification of 20 novel mutations and two novel polymorphisms in the FBN1 gene. *Hum. Mutat.*, **19**, 443–456.
18. Matyas,G., Arnold,E., Carrel,T., Baumgartner,D., Boileau,C., Berger,W. and Steinmann,B. (2006) Identification and in silico analyses of novel TGFBR1 and TGFBR2 mutations in Marfan syndrome-related disorders. *Hum. Mutat.*, **27**, 760–769.
19. Matyas,G., Alonso,S., Patrignani,A., Marti,M., Arnold,E., Magyar,I., Henggeler,C., Carrel,T., Steinmann,B. and Berger,W. (2007) Large genomic fibrillin-1 (FBN1) gene deletions provide

evidence for true haploinsufficiency in Marfan syndrome. *Hum. Genet.*, **122**, 23–32.
20. Meienberg,J., Rohrbach,M., Neuenschwander,S., Spanaus,K., Giunta,C., Alonso,S., Arnold,E., Henggeler,C., Regenass,S., Patrignani,A. *et al.* (2010) Hemizygous deletion of COL3A1, COL5A2, and MSTN causes a complex phenotype with aortic dissection: a lesson for and from true haploinsufficiency. *Eur. J. Hum. Genet.*, **18**, 1315–1321.
21. Klambauer,G., Schwarzbauer,K., Mayr,A., Clevert,D.A., Mitterecker,A., Bodenhofer,U. and Hochreiter,S. (2012) cn.MOPS: mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate. *Nucleic Acids Res.*, **40**, e69.
22. Fromer,M., Moran,J.L., Chambert,K., Banks,E., Bergen,S.E., Ruderfer,D.M., Handsaker,R.E., McCarroll,S.A., O'Donovan,M.C., Owen,M.J. *et al.* (2012) Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. *Am. J. Hum. Genet.*, **91**, 597–607.
23. Chen,K., Wallis,J.W., McLellan,M.D., Larson,D.E., Kalicki,J.M., Pohl,C.S., McGrath,S.D., Wendl,M.C., Zhang,Q., Locke,D.P. *et al.* (2009) BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat. Methods*, **6**, 677–681.
24. Benjamini,Y. and Speed,T.P. (2012) Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res.*, **40**, e72.
25. Zook,J.M., Chapman,B., Wang,J., Mittelman,D., Hofmann,O., Hide,W. and Salit,M. (2014) Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls. *Nat. Biotechnol.*, **32**, 246–251.
26. Meynert,A.M., Bicknell,L.S., Hurles,M.E., Jackson,A.P. and Taylor,M.S. (2013) Quantifying single nucleotide variant detection sensitivity in exome sequencing. *BMC Bioinformatics*, **14**, 195.
27. Alkan,C., Coe,B.P. and Eichler,E.E. (2011) Genome structural variation discovery and genotyping. *Nat. Rev. Genet.*, **12**, 363–376.
28. Wu,J., Grzeda,K.R., Stewart,C., Grubert,F., Urban,A.E., Snyder,M.P. and Marth,G.T. (2012) Copy number variation detection from 1000 genomes project exon capture sequencing data. *BMC Bioinformatics*, **13**, 305.
29. Okoniewski,M.J., Meienberg,J., Patrignani,A., Szabelska,A., Matyas,G. and Schlapbach,R. (2013) Precise breakpoint localization of large genomic deletions using PacBio and Illumina next-generation sequencers. *Biotechniques*, **54**, 98–100.
30. Neveling,K., Feenstra,I., Gilissen,C., Hoefsloot,L.H., Kamsteeg,E.J., Mensenkamp,A.R., Rodenburg,R.J., Yntema,H.G., Spruijt,L., Vermeer,S. *et al.* (2013) A post-hoc comparison of the utility of sanger sequencing and exome sequencing for the diagnosis of heterogeneous diseases. *Hum. Mutat.*, **34**, 1721–1726.
31. ENCODE Project Consortium. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.

Please refer to Appendix 2 for supplementary information.

### 2.1.2.2 Contribution of Authors

| | |
|---|---|
| **Janine Meienberg** | **Equally contributing first author; conceiving and planning the study, sequence capture and library preparation of the preliminary study, data analysis, writing of the manuscript** |
| Katja Zerjavic | Equally contributing first author; data analysis, writing of the manuscript |
| Irene Keller | Data analysis, editing of the manuscript |
| Michal Okoniewski | Data analysis, editing of the manuscript |
| Andrea Patrignani | Sequence capture and library preparation of the preliminary study, editing of the manuscript |
| Katja Ludin | Contribution to data analysis and editing of the manuscript |
| Zhenyu Xu | Contribution to data analysis and editing of the manuscript |
| Beat Steinmann | Contribution to data analysis and editing of the manuscript |
| Thierry Carrel | Contribution to data analysis and editing of the manuscript |
| Benno Roethlisberger | Contribution to data analysis and editing of the manuscript |
| Ralph Schlapbach | Contribution to data analysis and editing of the manuscript |
| Rémy Bruggmann | Data analysis, editing of the manuscript |
| Gabor Matyas | Initiation of the study, conceiving and planning the study, writing and editing of the manuscript |

# 2.2 Unpublished Results

## 2.2.1 aCGH Screening in Patients with Aortic Diseases

### 2.2.1.1 Introduction

Not only point mutations and INDELs but also larger deletions can play a role in AD [e.g. Matyas *et al*. 2007, Meienberg *et al*. 2010]. Routine mutation screening using Sanger sequencing and MLPA, however, may miss large deletions [Matyas *et al*. 2007, Shen and Wu 2009]. Currently, the method of choice for the detection of middle-sized and large deletions and duplications are microarrays. For the detection of genome-wide CNVs, microarrays are at the present faster, cheaper, and for data analysis less challenging than WGS or WES, for both of which the proof of concept for detection of large structural aberrations has been shown recently [Mills *et al*. 2011, Lonigro *et al*. 2011, Saintenac *et al*. 2011, Sathirapongsasuti *et al*. 2011, Koboldt *et al*. 2012].

In this thesis, AD patients with no mutation in genes covered by routine diagnostics for AD with unclear entity were screened for large deletions using microarrays. A custom-designed aCGH approach with arrays enriched for probes in exonic regions was used enabling the genome-wide detection of one-exon deletions. The goal of this approach was to assess the impact of large deletions in known AD candidate genes in our cohort as well as to identify new candidate genes for AD.

### 2.2.1.2 Material and Methods

#### Patients

65 unrelated patients with syndromic or non-syndromic AD were selected for this aCGH study. In this cohort, previous Sanger sequencing and MLPA analyses of *FBN1*, *TGFBR1*, and *TGFBR2* revealed no pathogenic sequence variant [Matyas *et al*. 2002, 2006, 2007]. In addition, we analysed 31 control samples, including such with known CNVs (9/31), known pathogenic mutation associated with AD (7/31) or unaffected individuals and family members (15/31). Data on clinical phenotypes were collected from medical records or during physical examinations. Informed consent was obtained from patients and family members and the study was approved by the responsible local ethics committee.

DNA was referred to us or extracted from EDTA-anticoagulated blood samples, saliva, tissue or cells cultured from aortic walls or skin biopsies (fibroblasts) using either QIAamp DNA Mini kit (Qiagen, Hilden, Germany) or Chemagic Magnetic Separation Module I (Chemagen, Perkin Elmer, Waltham, MA, USA) according to the manufacturers' instructions. DNA extracted from fibroblasts was treated with RNase prior to use for aCGH to reduce RNA contamination. In detail, ~3 µg genomic DNA (gDNA) was incubated with ribonuclease A

(10 µg/ml) and ribonuclease T1 (25 units/ml) at 37°C for 30 min in a total volume of 100 µl Tris-HCl buffer and subsequently purified using QIAamp DNA Mini kit (Qiagen, Hilden, Germany) according to manufacturers' instructions. DNA was quantified by OD measurements (NanoDrop, Thermo Scientific, Waltham, MA, USA).

**aCGH**

aCGH was performed using custom-designed 2.1/4.2 M NimbleGen aCGH (Roche Diagnostics, Risch, Switzerland) according to the manufacturers' instructions. These arrays contain the probes of commercially available 3×720 K/1.4 M genome-wide arrays (backbone) as well as 1.4/2.8 M additional custom-designed probes for exons to achieve higher resolution in these exonic regions. Briefly, 500 ng gDNA was labelled, quantified by OD measurements (NanoDrop, Thermo Scientific, Waltham, MA, USA), matched with control, and loaded on the array. Loaded arrays were hybridized for four days at 42°C and subsequently washed and scanned on a NimbleGen MS 200 Microarray Scanner (Roche Diagnostics, Risch, Switzerland). Data processing was performed using the software NimbleScan (Roche Diagnostics, Risch, Switzerland).

Data were analysed using Fast Adaptive States Segmentation Technique 2 (FASST2) segmentation with the significance threshold set to $5\times10^{-4}$, a maximal contiguous probe spacing of 1 Mb, and a minimal number of 3 probes per segment by the software Nexus Copy Number 7 (Biodiscovery, Hawthorne, CA, USA). We restricted the analysis to deletions, whereby the calling thresholds were set to log2 ratios of -0.5 and -1.1 for hemizygous and homozygous deletions, respectively. These settings proved to provide highest sensitivity and specificity during our evaluation process using our positive control samples.

**Confirmation and breakpoint analyses**

Promising deletions were confirmed using standard or LR-PCR followed by Sanger sequencing. Briefly, based on decreased microarray signal intensities, primers flanking the predicted deletion were designed and used in standard (expected fragment size <2-3 kb) or LR-PCR (Table 6). Accordingly, for standard PCR Hot Start thermostable DNA polymerase (HOT FIREPol; Solis BioDyne, Tartu, Estonia) was used whereas LR-PCR was performed using the Expand Long Template PCR System (Roche Diagnostics, Risch, Switzerland) according to manufacturers' instructions. Fragments containing breakpoints have been sequenced in both directions on an ABI PRISM 3730 Genetic Analyzer (Applied Biosystems, Zug, Switzerland) using PCR primers and the BigDye Terminator v1.1 cycle sequencing kit (Applied Biosystems, Zug, Switzerland).

**Table 6**. Primers designed for PCR and/or long-range PCR according to the results of the microarray analyses.

| | Forward Primer | | Reverse Primer | | Unclear | |
|---|---|---|---|---|---|---|
| Patient | Name | Sequence (5'→3') | Name | Sequence (5'→3') | Region* | Assay |
| 420 | N328_S | GGGAAGAGGGGCAAAACTATC | N4552_AS | CCTGCCCCCTCCTCTCTC | 3,951 bp | LR-PCR |
| 30544 | N303_F2 | GGAGGGTCAGGACAGGAAGTG | N3144_R | CTGGCTACCTTCCCTTGTCAAC | 15,004 bp | LR-PCR |
| 127 | N253_F | TCCACGTAGTTTTTGTTCTTTCAGTT | N2507_R | GGTTTTCTCCAGTGCCAGTTC | 1,850 bp | PCR |
| 361 | N418_F | AGGAAGTGAAGGGTTCTCTTTTCTAT | N3410_R | ACCTGTTGCCTGCCTGAT | 2,523 bp | PCR |
| 37 | N439_F | TGCTCACCAATCAGGGAATG | N2501_R | GGCACCCCAATCCGTATG | 1,949 bp | PCR |
| 303 | N69_F | AGCCATAAAAGGAACGACATCA | N4182_R | AATCCCCTGCTTCAATAAATCA | 3,605 bp | LR-PCR |
| 30576 | N253_F | CCATTCTCACAGGTTTCCAGTC | N3063_R | GCCTCCCTCTCCAAACATTC | 2,505 bp | LR-PCR |
| 100 | N374_F | GCTTTCGGAGAACAGAACACTTG | N5380_R | AGGGGTGGGGTTGTGGA | 4,741 bp | LR-PCR |
| 31778 | N398_F | GCACCACTCTCTCCACTGACAA | N6768_R | AGATGTCCCTGCCACCTGAAT | 6,249 bp | LR-PCR |
| 110 | N428_F | GGGTGGGGACAAACACTTCTC | N25697_R | TATTACTGGGTGGGTCTGATGTGA | 25,173 bp | LR-PCR |
| 350 | N456_F | ACCAGAAGGAGCGAGAGATTATG | N21208_R | GGTGCTTTTGGGTTTTCTGTGTA | 20,619 bp | LR-PCR |

*Regions between last normal and first reduced array signal, in which breakpoints are expected; LR-PCR: long-range PCR; PCR; standard PCR.

## MLPA

If available, deletions in promising AD candidate genes were confirmed by MLPA in the index patients as well as in their relatives. The MLPA kit P231-A2 (MRC-Holland, Amsterdam, the Netherlands) contains probes for 9 of the 18 exons of *FGFR2* (NM_00141.4) including one of the two exons deleted in patient 420 (exon 3). Briefly, MLPA was performed using 100 ng template DNA according to the manufacturers' instructions. MLPA fragments were separated by capillary electrophoresis on an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems, Zug, Switzerland). Each MLPA signal was normalized and compared to the corresponding peak area obtained in control DNA samples. Deviations >30% were suspected as alterations and verified by repeated MLPA analysis if not already known from aCGH.

## NGS data

In order to assess the role of promising AD candidate genes identified by aCGH in our cohort of AD patients with unknown molecular basis, these genes were screened for mutations in our NGS data set. This NGS data set includes WGS with 60× coverage for 101 AD index patients, which were sequenced on a HiSeq X Ten system (Illumina, San Diego, CA, USA) by an external vendor using 2 µg gDNA, either Illumina's TruSeq Nano DNA Sample Preparation Kit, which is not PCR-free (43 patients), or Illumina's TruSeq PCR-Free Sample Preparation Kit (58 patients), and 2 × 150 bp paired-end sequencing according to the manufacturers' instructions for 350-bp insert size. In addition, there are also NGS data from clinically focused exome sequencing including exons and flanking intronic sequences of ~7,600 clinically relevant genes for 13 index patients sequenced at 60× coverage by an external vendor on a HiSeq2500 system (Illumina, San Diego, CA, USA) using 4 µg gDNA, the Accuracy and Content Enhanced (ACE) clinical exome enrichment platform (ACEv2, Personalis, Inc., Menlo Park, CA, USA), and 2 × 100 bp paired-end sequencing according to their standard workflow. For already known AD candidate genes additional 15 patients were available for mutation screening, which were sequenced by an external vendor at 300× coverage on a MiSeq system (Illumina, San Diego, CA, USA) using 2 µg gDNA, a SureSelect

custom-designed enrichment panel (SureSelect Protocol Version 1.2, Agilent, Santa Clara, CA, USA) including exons and intronic flanking sequences for our list of AD candidate genes (Appendix 3), and 2 × 150 bp paired-end sequencing according to their standard workflow. Data analysis was performed by displaying coverage tracks of provided BAM files in Alamut visual (Interactive Biosoftware, Rouen, France) and screening coding exons and 50 bp flanking intronic sequences of genes of interest for nucleotide changes with at least 20% non-reference allele fraction or reduced coverage due to small deletions.

Patients with deletions affecting a possible AD candidate gene were sequenced using the AD candidate gene panel except for patients 420 and 30576, for which WGS data are available. Using the generated NGS data, we screened these patients for further possibly disease-causing or modifying mutations in our AD candidate genes. In detail, we filtered the annotated results of the AD candidate gene panel analysis provided by the external vendor for sequence variants with at least 20% mutant allele and a population frequency of less than 1% according to dbSNP release 137 (http://www.ncbi.nlm.nih.gov/SNP), which are non-synonymous and predicted to be deleterious by at least one of the prediction tools SIFT [Kumar *et al*. 2009], PolyPhen [Ramensky *et al*. 2002], and Condel [Gonzalez-Perez and Lopez-Bigas 2011], lead to a frameshift or affect a splice site. For patients 420 and 30576, we used the platform knoSYS v3.1.02 (Knome, Waltham, MA, USA) for annotation and filtering of the sequence variants. We restricted the analysis also to the same set of AD candidate genes and used corresponding filtering criteria as for the other patients.

## Exon-by-exon Sanger sequencing for *B3GLCT*

**Table 7**. Primers used for gDNA sequencing of *B3GLCT*.

| | Forward Primer | | Reverse Primer | | Fragment |
|---|---|---|---|---|---|
| Exon | Name | Sequence (5'→3') | Name | Sequence (5'→3') | length |
| Exon 1 | N58F | CTGGAGGGGGCAGAGGTCAGA | N605R | GAGAGGGCGCTGCTGACGG | 566 bp |
| Exon 2 | N62F | TGAGCAAAATATGTCTTGTT | N348R | ATCCATGTTGACCACATTAT | 306 bp |
| Exon 3 | N5F | CTCCGTGGTCCCTTAGGT | N373R | TTTGCCCCACATCTTCAG | 386 bp |
| Exon 4 | N59F | TTTTCCCATTAAGAGTTTACTG | N396R | CAAAATGAGTCAGAGGAGTTAT | 359 bp |
| Exon 5 | N8F | GGAAAAAGAAGTGGTTTGAA | N328R | GAAAACCACACCCTCTTAATA | 341 bp |
| Exon 6 | N91F | CCCTTCATTCACTTCCTACTG | N448R | TAAGCTACAAACTTCAAAGAGCA | 380 bp |
| Exon 7 | N58F | TCTGTGCTAATAACTCTTTATCACC | N410R | CTGGCTGAAAAACTCATTTG | 372 bp |
| Exon 8 | N37F | TTCCCTCAAGGAATTTAAAC | N364R | CTGGAGTGCTATGAAATTATCT | 349 bp |
| Exon 9 | N93F | TTCTGCTTTCCCTTGAGATA | N460R | TATTTGGGTGACAGAATCAG | 387 bp |
| Exon 10 | N66F | TGTTTGGTATCCTTGACATT | N347R | AACATCCGAATTGTCATTATC | 302 bp |
| Exon 11 | N1F | AAGAGGGTTTTAGTCAACAA | N409R | GAGAGAAGGGAAAAAACTAA | 428 bp |
| Exon 12 | N66F | AGCTGCATTTTGGCATGTAA | N401R | GTTCCTTAAGAGGATTGGTTCA | 357 bp |
| Exon 13 | N100F | GTGGGATGTAAGAACCATAAA | N401R | CAACCACCATTCACAGAGT | 320 bp |
| Exon 14 | N4F | AAGGGCCAGAAGACTAAAA | N454R | AGTCCTGGTTAAGGACATTC | 470 bp |
| Exon 15 | N2F | AGCAAAGCACCTGGAGTTAG | N657R | ACAGGAAAGGGACTTCTGGTA | 676 bp |

All 15 exons and flanking intronic sequences of *B3GLCT* (NM_194318.3), which is associated with an autosomal recessive disorder, were sequenced in patient 30544, who has a heterozygous deletion in this gene, in order to assess whether there could be a compound heterozygous situation. In detail, PCR was performed using 10 ng gDNA, Hot Start thermostable DNA polymerase (HOT FIREPol; Solis BioDyne, Tartu, Estonia), and primers

listed in Table 7. The amplicons were sequenced in both directions on an ABI PRISM 3100/3730 Genetic Analyzer (Applied Biosystems, Zug, Switzerland) using PCR primers and the BigDye Terminator v1.1 cycle sequencing kit (Applied Biosystems, Zug, Switzerland).

### Statistical analyses

For proportions, the upper and lower limits of the 95% confidence interval were calculated [Matyas *et al*. 2002].

## 2.2.1.3 Results

**Table 8**. Deletions selected for confirmation.

| Patient | Gene | Effect | Location | Size of Deletion | Current State | Comment | Clinical Significance |
|---|---|---|---|---|---|---|---|
| **420** | ***FGFR2*** | **Ex3-4 (in frame)** | **10q26.13** | **22.6 kb** | **Characterized** | **AD candidate gene** | **Probable** |
| 30544 | *B3GLCT* | Ex6 (frameshift) | 13q12.3 | 10.7 kb | Characterized | --- | Possible |
| 127 | *PCDHGA8, PCDHGB4, PCDHGB5* | In all 3 genes Ex1 affected | 5q31.3 | 10.2 kb | Characterized | --- | Possible |
| 361 | *VWA3A* | Ex26 (frameshift) | 16p12.2 | 1.1 kb | Characterized | --- | Possible |
| 37 | *PCDHGA11* | 3'end of Ex1 | 5q31.3 | 4 kb | Characterized | Also in healthy mother | Improbable |
| 303 | *MMP26, OR51L1* | Complete genes | 11p15.4 | ~29-32 kb | Confirmed | --- | Further analyses needed |
| 30576 | *ITGAE* | Ex23-26 (in frame) | 17p13.2 | ~3-6 kb | Confirmed | --- | Further analyses needed |
| 100 | *NDUFA6* | Ex2-3 (contains stop codon) | 22q13.2 | ~2-6 kb | Confirmed | --- | Further analyses needed |
| 31778 | *MYO19* | Ex4-5 (in frame) | 17q12 | ~1-7 kb | Confirmed | --- | Further analyses needed |
| 110 | *CNTNAP2* | Ex2-3 (frameshift) | 7q35 | ~350-380 kb | Ongoing | --- | Further analyses needed |
| 350 | *COL6A5* | 5'UTR | 3q22.1 | ~2-22 kb | Ongoing | --- | Further analyses needed |

Ex, exon; UTR, untranslated region; characterized, deletion breakpoints identified; confirmed, deletion specific fragment amplified by LR-PCR but breakpoints are outside the region covered by Sanger sequencing; ongoing, confirmation using LR-PCR is still ongoing; bold, already selected as potential AD candidate gene from study of literature (Appendix 3).

Our aCGH analysis revealed 11 deletions detected only once in the cohort of 65 AD patients and in none of the 31 control samples, which affect potential AD candidate genes (Table 8). One of these deletions affects a gene (*FGFR2*) in our list of possible AD candidate genes (Appendix 3). This deletion truncating *FGFR2* was confirmed in patient 420 using LR-PCR and Sanger sequencing (Figure 12). *FGFR2*, which encodes fibroblast growth factor receptor 2, is associated with a broad spectrum of autosomal dominant phenotypes [Wilkie 2005]. This deletion at Chr10:123,312,993-123,335,592 (most telomeric position of four possible) has a length of 22,600 bp, includes complete exon 3 and exon 4 of *FGFR2* (NM_00141.4), and leads to an in frame deletion of 115 amino acids (p.E37_R152delinsG). Four disease-causing mutations have been described for this region in the Human Gene Mutation database (HGMD, https://portal.biobase-international.com/hgmd/pro/start.php), which are associated with Crouzon syndrome, craniosynostosis, and cleft lip and palate, respectively. Since *FGFR2* activates ERK [Hadari *et al*. 2001], which is also activated by non-canonical TGFβ signalling (1.1.3.1, Figure 7) and known to be involved in aneurysm development [Holm *et al*. 2011], this gene is in our list of candidate genes for AD (Appendix 3). Patient 420 has an affected aorta and positive family history. Unfortunately, the deletion does not segregate with the phenotype within the family. This suggests that there must be a mutation in another gene and that this deletion has at most modifying effects. However, no promising

sequence variant has been detected in our AD candidate genes in patient 420 (Table 9). The suspected rather low impact of this gene on AD was confirmed by screening 24 additional patients with MLPA for deletions in this gene and analysis of our NGS data set including 129 index patients as no further deletion and no novel clearly pathogenic mutation was detected (Table 10).

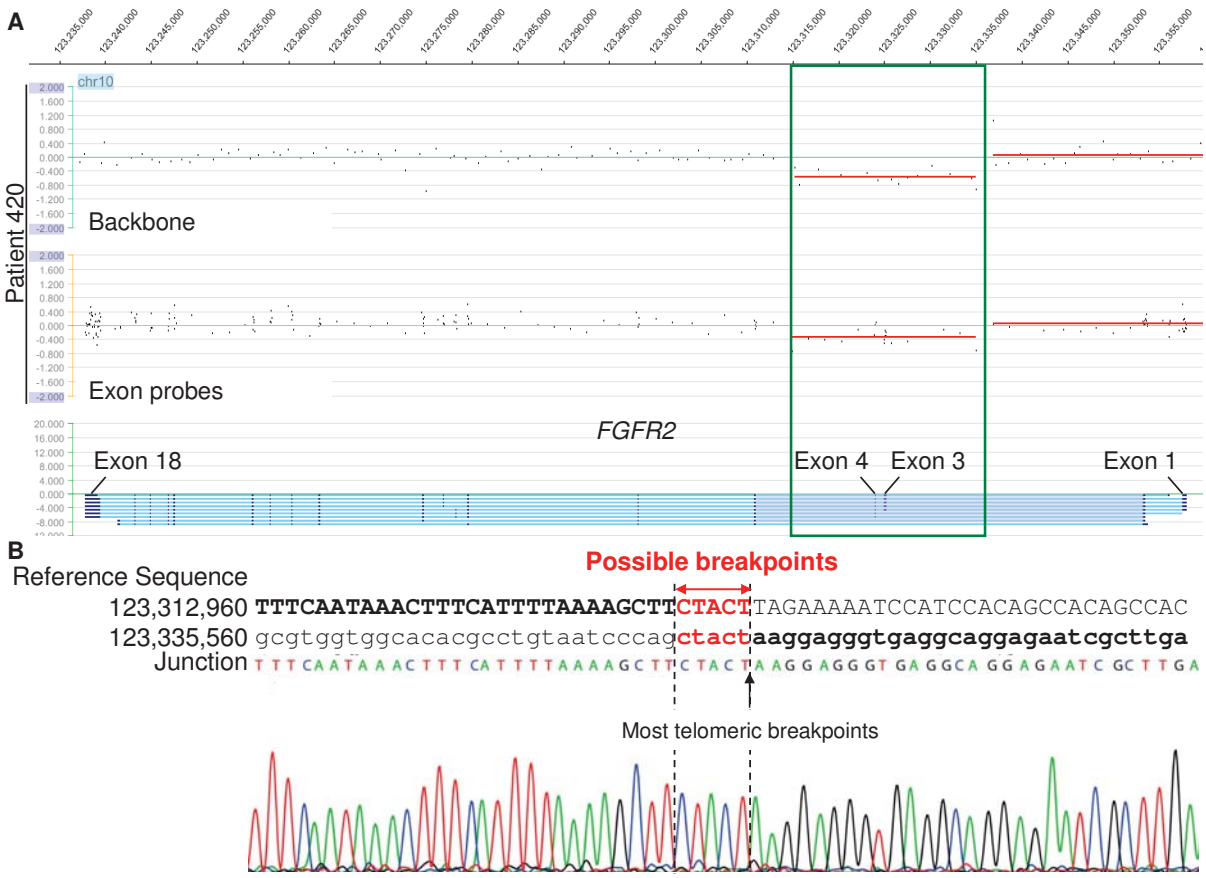



**Figure 12.** Deletion affecting *FGFR2*. **A**: aCGH signal of patient 420 displayed in SignalMap (Roche Diagnostics, Risch, Switzerland). **B**: Sanger sequences spanning the breakpoints.

**Table 9**. Additional potentially disease-causing sequence variants detected in patients with promising deletion.

| Patient | Gene | NM_Number | cDNA | Protein | ExAC (hom)* | PhyloP | PhastCons | HGMD/ClinVar | SIFT | MutationTaster | Our NGS Data |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 100 | *LTBP4* | NM_001042544.1 | c.4499A>T (hom) | p.Tyr1500Phe | 167/14306 (0) | N/A | 1.00 | N/A | Tolerated (score: 0.35) | Disease causing (p-value: 0.869) | 9× het |
| **100** | ***NOTCH1*** | **NM_017617.3** | **c.3125_3126delinsTT** | **p.Ser1042Phe** | **---** | **3.68/0.77** | **1.00/1.00** | **not found** | **Deleterious (score: 0)** | **Disease causing (p-value: 0.731)** | **not found** |
| **110** | ***NOTCH3*** | **NM_000435.2** | **c.3466C>T** | **p.Leu1156Phe** | **---** | **3.43** | **1.00** | **not found** | **Deleterious (score: 0)** | **Disease causing (p-value: 1)** | **not found** |
| 127 | *APC* | NM_001127510.2 | c.3386T>C | p.Leu1129Ser | 195/120606 (2) | 2.38 | 1.00 | HGMD: DM?<br>ClinVar: benign | Deleterious (score: 0.02) | Disease causing (p-value: 0.821) | 1× het |
| 127 | *CHD7* | NM_017780.3 | c.8950C>T | p.Leu2984Phe | 543/84954 (3) | 1.90 | 0.66 | ClinVar: benign | Deleterious (score: 0) | Polymorphism (p-value: 0.914) | 1× het |
| **127** | ***HAS1*** | **NM_001523.2** | **c.307_308delinsTT** | **p.Ala103Phe** | **---** | **4.73/4.73** | **0.99/1.00** | **not found** | **---** | **---** | **not found** |
| **127** | ***LEFTY2*** | **NM_003240.3** | **c.74A>T** | **p.Glu25Val** | **---** | **2.55** | **0.99** | **not found** | **Deleterious (score: 0.04)** | **Disease causing (p-value: 0.992)** | **not found** |
| **127** | ***NPHP3*** | **NM_153240.4** | **c.3610G>A** | **p.Val1204Ile** | **---** | **3.92** | **1.00** | **not found** | **Tolerated (score: 0.1)** | **Disease causing (p-value: 1)** | **not found** |
| 127 | *TGFBR1* | NM_004612.2 | c.76_78dup | p.Ala26dup | N/A | --- | --- | not found | --- | --- | N/A |
| 127 | *THBS2* | NM_003247.3 | c.2323C>A | p.Arg775Ser | 1/120874 (0) | 5.69 | 1.00 | not found | Deleterious (score: 0) | Disease causing (p-value: 1) | not found |
| **127** | ***TSC2*** | **NM_000548.3** | **c.1375G>T** | **p.Gly459Cys** | **---** | **3.11** | **1.00** | **not found** | **Deleterious (score: 0.04)** | **Polymorphism (p-value: 0.669)** | **not found** |
| 303 | *LTBP3* | NM_001130144.2 | c.3281C>T | p.Pro1094Leu | 1/9640 (0) | 1.09 | 0.93 | not found | Tolerated (score: 0.12) | Disease causing (p-value: 0.836) | 1× het |
| 303 | *NOS1* | NM_001204218.1 | c.721G>A | p.Asp241Asn | 242/116436 (1) | 3.51 | 1.00 | not found | Tolerated (score: 0.11) | Disease causing (p-value: 0.993) | 1× het |
| 350 | *HSPG2* | NM_001291860.1 | c.3548C>T | p.Thr1183Met | 5/117968 (0) | 5.29 | 1.00 | not found | Deleterious (score: 0.03) | Polymorphism (p-value: 0.943) | not found |
| 350 | *KMT2D* | NM_003482.3 | c.7670C>T | p.Pro2557Leu | 1004/118200 (11) | 1.25 | 0.99 | ClinVar: benign | Tolerated (score: 0.37) | Disease causing (p-value: 0.953) | 1× het |
| **350** | ***NOS1*** | **NM_001204218.1** | **c.2989A>G** | **p.Ile997Val** | **---** | **2.87** | **1.00** | **not found** | **Deleterious (score: 0)** | **Disease causing (p-value: 1)** | **not found** |
| 350 | *RAG1* | NM_000448.2 | c.1346G>A | p.Arg449Lys | 1266/121268 (13) | 1.58 | 0.96 | HGMD: DM? | Tolerated (score: 0.22) | Disease causing (p-value: 0.945) | 4× het |
| 350 | *RAG2* | NM_000536.3 | c.1158C>A | p.Phe386Leu | 1244/121370 (12) | 1.66 | 1.00 | ClinVar: benign | Deleterious (score: 0) | Disease causing (p-value: 1) | 4× het |
| 361 | *KMT2D* | NM_003482.3 | c.2420C>A | p.Ser807Tyr | 2/105566 (0) | 2.14 | 1.00 | not found | Deleterious (score: 0) | Polymorphism (p-value: 0.936) | not found |
| 361 | *VCAN* | NM_004385.4 | c.574G>A | p.Gly192Arg | 219/121396 (1) | 6.26 | 1.00 | not found | Deleterious (score: 0) | Disease causing (p-value: 1) | 4× het |
| 420 | *ARVCF* | NM_001670.2 | c.1822C>T | p.Arg608Cys | 145/113310 (2) | 2.95 | 1.00 | not found | Deleterious (score: 0.01) | Disease causing (p-value: 1) | 1× het |
| 420 | *CDKN2B* | NM_004936.3 | c.256G>A | p.Asp86Asn | 166/112740 | 4.56 | 1.00 | HGMD: DM? | Deleterious (score: 0) | Disease causing (p-value: 1) | not found |
| 420 | *COL9A3* | NM_001853.3 | c.1698_1706del | p.Pro567_Gly569del | 24/118584 (0) | --- | --- | not found | --- | --- | not found |
| 420 | *HOXA1* | NM_005522.4 | c.972delG | p.Ser325Leufs*7 | 10/121124 (0) | -0.28 | 0.92 | not found | --- | --- | not found |
| 420 | *LTBP4* | NM_001042544.1 | c.4796C>T | p.Ser1599Phe | 140/115504 (0) | 2.14 | 1.00 | N/A | Deleterious (score: 0) | --- | 1× het |
| 420 | *THSD4* | NM_024817.2 | c.1412G>A | p.Arg471Gln | 2/120684 (0) | 3.51 | 1.00 | not found | Deleterious (score: 0) | Disease causing (p-value: 0.971) | not found |
| 30544 | *NOTCH1* | NM_017617.3 | c.1862G>A | p.Arg621His | 138/115136 (0) | 4.00 | 0.99 | ClinVar: N/A | Deleterious (score: 0) | Disease causing (p-value: 1) | 2× het |
| 30576 | *NOTCH2* | NM_024408.2 | c.7223T>A | p.Leu2408His | 217/121356 | 1.05 | 1.01 | ClinVar: N/A | Deleterious (score: 0) | Disease causing (p-value: 0.739) | 1× het |
| **30576** | ***PKD1*** | **NM_001009944.2** | **c.11854C>T** | **p.Arg3952Cys** | **---** | **1.90** | **0.98** | **Nein** | **Tolerated (score: 0.18)** | **Disease causing (p-value: 0.915)** | **not found** |
| 30576 | *SMAD7* | NM_005904.3 | c.115G>A | p.Gly39Arg | 34/10402 (0) | 3.11 | 1.00 | Nein | Deleterious (score: 0) | Disease causing (p-value: 0.999) | 2× het |
| 30576 | *TGFBR3* | NM_003243.4 | c.2329C>T | p.Pro777Ser | 805/121368 (0) | 3.03 | 1.00 | HGMD: DM | Tolerated (score: 0.39) | Disease causing (p-value: 0.876) | 4× het |

Information on sequence variants taken from Alamut visual (Interactive Biosoftware, Rouen, France); *number of alternative alleles/total sequenced alleles (number of homozygous individuals); N/A, not available; --- no information; het, heterozygous; hom, homozygous; DM, disease causing mutation; DM?, likely disease causing mutation; ExAC, Exome Aggregation Consortium browser (http://exac.broadinstitute.org); PhiloP, phylogenetic P-values [Pollard *et al.* 2010]; PhastCons, PHylogenetic Analysis with Space/Time models Conservation [Siepel *et al.* 2005]; HGMD, Human Gene Mutation Database (https://portal.biobase-international.com/hgmd/pro/start.php); ClinVar, http://www.ncbi.nlm.nih.gov/clinvar [Landrum *et al.* 2014]; SIFT, Sorting Intolerant From Tolerant (http://sift.jcvi.org) [Kumar *et al.* 2009]; MutationTaster, http://www.mutationtaster.org [Schwarz *et al.* 2014]; bold, hot/promising sequence variants, for which further analyses are needed.

**Table 10**. Potentially pathogenic sequence variants in potential AD candidate genes identified by aCGH detected in our NGS data set.

| Gene | NM_number | cDNA | Protein | ExAC (hom)* | phyloP | phastCons | HGMD/ClinVar | SIFT | MutationTaster | Our NGS Data |
|---|---|---|---|---|---|---|---|---|---|---|
| *B3GLCT* | NM_194318.3 | c.1405G>T | p.Asp469Tyr | 13/121404 (0) | 4.56 | 1 | not found | Deleterious (score: 0) | Disease causing (p-value: 1) | 1× het |
| *CNTNAP2* | NM_014141.5 | c.436G>A | p.Val146Ile | 6/121392 (0) | 5.69 | 1 | not found | Deleterious (score: 0) | Disease causing (p-value: 0.994) | 1× het |
| *CNTNAP2* | NM_014141.5 | c.854G>C | p.Gly285Ala | 640/121374 (1) | 5.94 | 1 | ClinVar: US | Tolerated (score: 0.61) | Disease causing (p-value: 0.998) | 2× het |
| *CNTNAP2* | NM_014141.5 | c.2651G>A | p.Arg884Gln | 29/121408 (0) | 2.95 | 1 | not found | Deleterious (score: 0) | Disease causing (p-value: 0.999) | 1× het |
| ***CNTNAP2*** | **NM_014141.5** | **c.2782G>A** | **p.Gly928Arg** | **--** | **3.03** | **0.98** | **not found** | **Deleterious (score: 0.01)** | **Disease causing (p-value: 1)** | **1× het** |
| *COL6A5* | NM_001278298.1 | c.297_299del | p.Lys99del | 0.0086%*** | --- | --- | not found | --- | --- | 1× het |
| *COL6A5* | NM_001278298.1 | c.1742C>T | p.Ala581Val | 95/19742 (2) | 4 | 0.99 | not found | Deleterious (score: 0) | Disease causing (p-value: 0.974) | 1× het |
| *COL6A5* | NM_001278298.1 | c.3598G>A | p.Asp1200Asn | 7/17904 (0) | 4.81 | 1 | not found | Deleterious (score: 0.03) | Disease causing (p-value: 0.975) | 1× het |
| ***COL6A5*** | **NM_001278298.1** | **c.4202G>C** | **p.Gly1401Ala** | **---** | **3.76** | **0.79** | **not found** | **Deleterious (score: 0)** | **Disease causing (p-value: 0.607)** | **1× het** |
| *COL6A5* | NM_001278298.1 | c.4624A>C | p.Lys1542Gln | --- | 0.45 | 0.87 | not found | Deleterious (score: 0.04) | Polymorphism (p-value: 1) | 1× het |
| *COL6A5* | NM_001278298.1 | c.4868C>A | p.Pro1623His | 575/19786 (15) | 3.11 | 0.99 | not found | Deleterious (score: 0.03) | Polymorphism (p-value: 0.996) | 2× het |
| ***COL6A5*** | **NM_001278298.1** | **c.4987C>T** | **p.Arg1663*** | **---** | **0.12** | **0.19** | **not found** | **---** | **---** | **1× het** |
| *COL6A5* | NM_001278298.1 | c.5423C>T | p.Ser1808Leu | 11/19812 (0) | 0.61 | 0 | not found | Deleterious (score: 0) | Polymorphism (p-value: 1) | 3× het |
| ***COL6A5*** | **NM_001278298.1** | **c.6814G>T** | **p.Glu2272*** | **849/88130 (11)** | **0.93** | **0.99** | **not found** | **---** | **---** | **1× het** |
| *FGFR2* | NM_000141.4 | c.1085-41G>A | ---** | 98/121038 (0) | 2.14 | 1 | not found | --- | --- | 1× het |
| *ITGAE* | NM_002208.4 | c.34+2T>G | ---** | 3/120774 (0) | 2.14 | 0.99 | not found | --- | --- | 1× het |
| *ITGAE* | NM_002208.4 | c.3237+25A>G | ---** | 84/121146 (0) | 0.61 | 0.05 | not found | --- | --- | 1× het |
| *MYO19* | NM_001163735.1 | c.721-6C>G | ---** | 350/107020 (3) | 1.01 | 0.1 | not found | --- | --- | 1× het |
| *MYO19* | NM_001163735.1 | c.1188C>T | p.Asn396Asn** | --- | 1.98 | 1 | not found | --- | --- | 1× het |
| *NDUFA6* | NM_002490.3 | c.400C>T | p.His134Tyr | 341/121408 (1) | 3.35 | 1 | not found | Tolerated (score: 0.3) | Disease causing (p-value: 0.999) | 2× het |
| *PCDHGA8* | NM_032088.1 | c.193C>T | p.Arg65Cys | 282/120336 (3) | 4.08 | 1 | N/A | Deleterious (score: 0) | Disease causing (p-value: 0.999) | 1× het |
| *PCDHGA8* | NM_032088.1 | c.329T>A | p.Val110Asp | --- | 1.66 | 0.23 | N/A | Deleterious (score: 0) | Polymorphism (p-value: 1) | 1× het |
| *PCDHGA8* | NM_032088.1 | c.931G>A | p.Glu311Lys | 2/120454 (0) | 1.09 | 0.98 | N/A | Deleterious (score: 0.04) | Polymorphism (p-value: 0.991) | 1× het |
| *PCDHGA8* | NM_032088.1 | c.1432G>A | p.Asp478Asn | 257/120484 (0) | 5.77 | 1 | N/A | Deleterious (score: 0) | Disease causing (p-value: 0.997) | 1× het |
| *PCDHGA8* | NM_032088.1 | c.1508C>G | p.Ser503Trp | 3402/120640 (91) | 5.61 | 1 | N/A | Deleterious (score: 0) | Polymorphism (p-value: 0) | 2× het |
| *PCDHGA11* | NM_018914.2 | c.1405G>T | p.Gly469Cys | 217/118980 (1) | 4.64 | 1 | N/A | Deleterious (score: 0.01) | Disease causing (p-value: 0.863) | 1× het |
| *PCDHGA11* | NM_018914.2 | c.1883G>A | p.Arg628His | 1078/120328 (11) | 2.55 | 1 | N/A | Deleterious (score: 0) | Polymorphism (p-value: 0.993) | 2× het |
| *PCDHGA11* | NM_018914.2 | c.2061_2063del | p.Ser688del | 1013/120768 (7) | --- | --- | N/A | --- | --- | 2× het |
| *PCDHGB4* | NM_003736.2 | c.466T>C | p.Ser156Pro | --- | -0.44 | 0 | N/A | --- | --- | 1× het |
| *PCDHGB4* | NM_003736.2 | c.565A>G | p.Ser189Gly | --- | 0.37 | 0 | N/A | --- | --- | 1× het |
| *PCDHGB4* | NM_003736.2 | c.681G>C | p.Gln227His | 236/120334 | -0.12 | 0 | N/A | --- | --- | 1× het |
| *PCDHGB4* | NM_003736.2 | c.1225G>A | p.Asp409Asn | 46/120752 (0) | 4.24 | 1 | N/A | --- | --- | 1× het |
| *PCDHGB4* | NM_003736.2 | c.1393C>T | p.Pro465Ser | 539/120750 (0) | 2.14 | 0.69 | N/A | --- | --- | 1× het |
| *PCDHGB5* | NM_018925.2 | c.565A>G | p.Ser189Gly | 192/119742 (0) | 0.37 | 0 | N/A | --- | --- | 2× het |
| *PCDHGB5* | NM_018925.2 | c.886G>A | p.Glu296Lys | 1390/119636 (15) | 0.85 | 0.61 | N/A | --- | --- | 2× het |
| *PCDHGB5* | NM_018925.2 | c.1562T>C | p.Leu521Pro | 48/119310 (0) | 1.34 | 1 | N/A | --- | --- | 1× het |
| *PCDHGB5* | NM_018925.2 | c.1669C>T | p.Arg557Trp | 1327/118630 (9) | -0.36 | 0 | N/A | --- | --- | 2× het |
| *PCDHGB5* | NM_018925.2 | c.1781C>T | p.Ser594Leu | 1034/118774 (6) | 3.03 | 0.99 | N/A | --- | --- | 2× het |
| *VWA3A* | NM_173615.3 | c.1720G>A | p.Ala574Thr | 2/117734 (0) | 3.76 | 1 | N/A | Deleterious (score: 0) | Disease causing (p-value: 1) | 1× het |
| *VWA3A* | NM_173615.3 | c.1744C>T | p.Ala582Trp | 5/114356 (0) | 0.29 | 0.44 | N/A | Deleterious (score: 0.01) | Polymorphism (p-value: 1) | 1× het |
| *VWA3A* | NM_173615.3 | c.1936C>A | p.Leu646Ile | 223/120424 (1) | 2.47 | 1 | N/A | Tolerated (score: 0.06) | Disease causing (p-value: 0.998) | 2× het |
| *VWA3A* | NM_173615.3 | c.1982G>A | p.Arg661His | 339/119108 (2) | 3.43 | 1 | N/A | Tolerated (score: 0.2) | Disease causing (p-value: 0.75) | 1× het |
| *VWA3A* | NM_173615.3 | c.2891T>G | p.Leu964Arg | 3/57440 (1) | 3.27 | 1 | N/A | Deleterious (score: 0) | Disease causing (p-value: 0.951) | 1× het |
| *VWA3A* | NM_173615.3 | c.2897C>A | p.Thr966Lys | --- | 0.77 | 0.67 | N/A | Deleterious (score: 0) | Polymorphism (p-value: 0.945) | 1× het |
| *VWA3A* | NM_173615.3 | c.3325G>A | p.Gly1109Arg | 100/101682 (0) | 3.68 | 0.94 | N/A | Deleterious (score: 0) | Disease causing (p-value: 1) | 2× het |

Information on sequence variants taken from Alamut visual (Interactive Biosoftware, Rouen, France); *number of alternative alleles/total sequenced alleles (number of homozygous individuals); ** effect on splicing predicted; ***detailed information not available; bold, hot/promising sequence variants, for which further analyses are needed; US, uncertain significance; details to further abbreviations are as listed in Table 9.

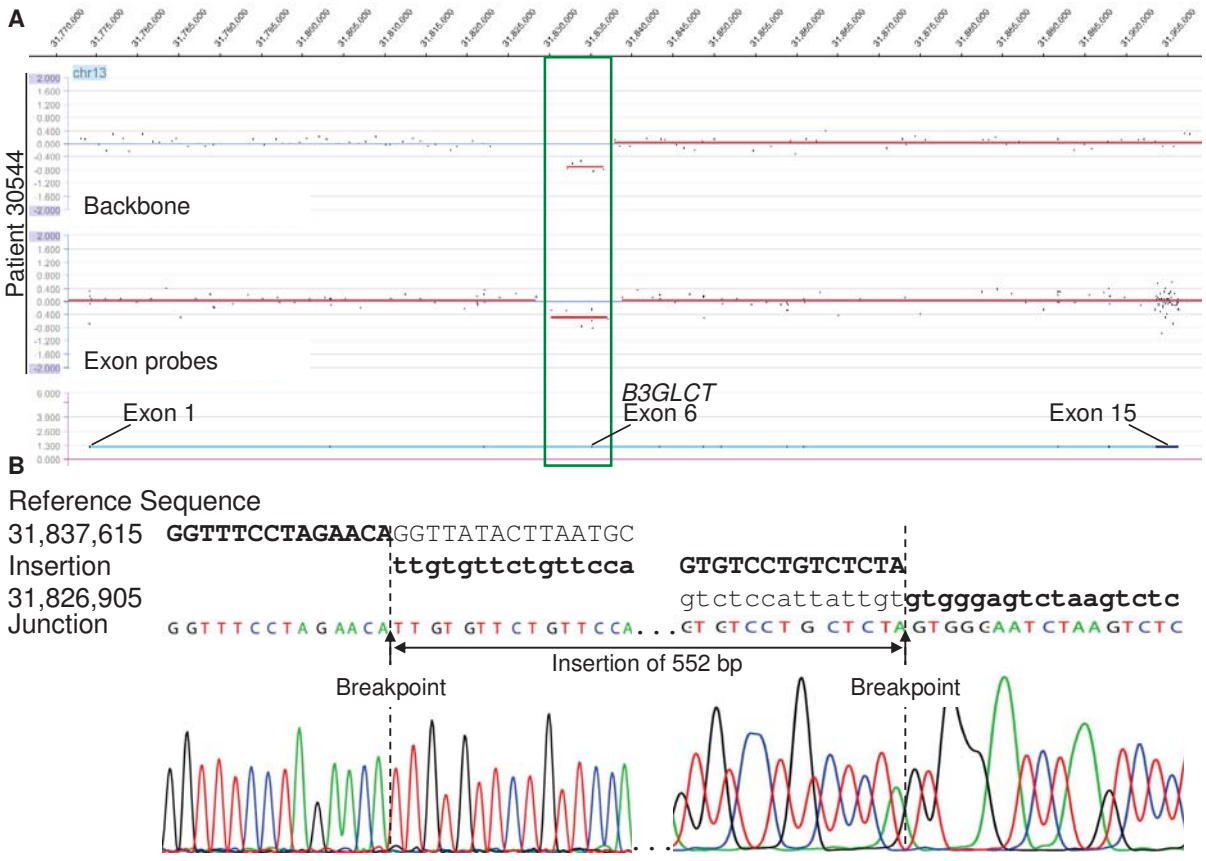



**Figure 13.** Deletion affecting *B3GLCT*. **A**: aCGH signal of patient 30544 displayed in SignalMap (Roche Diagnostics, Risch, Switzerland). **B**: Sanger sequences spanning the breakpoints.

Patient 30544 has a deletion in the gene *B3GLCT* which is associated with autosomal recessive Peters-Plus syndrome, a congenital disorder of glycosylation [Lesnik Oberstein *et al*. 2006]. This deletion of 10,710 bp at chr13:31,826,892-31,837,601 affects complete exon 6 of *B3GLCT* (Figure 13) and leads to a frameshift and premature stop codon. Consequently, degradation of the transcript by nonsense-mediated mRNA decay (NMD) leading to functional haploinsufficiency may be expected, but has not yet been confirmed. This rather complex SV contains an insertion of 552 bp, which consists of 129 bp of the deleted region, 17 bp in the correct orientation and 112 bp inverted, respectively, as well as 423 bp of unknown origin, between the two breakpoints. The characteristic feature of patients with Peters-Plus syndrome is a specific malformation in the eye known as Peters' anomaly. In addition, these patients present with short stature including short limbs and may have cleft lip/palate, hypertelorism, and defects in the central nervous system, heart, and various other organs. *B3GLCT* encodes beta-1,3-glucosyltransferase, which is responsible for a disaccharide modification specific to thrombospondin type I repeat domains and was demonstrated to be present in thrombospondin I, properdin, F-spondin, a disintegrin and metalloproteinase with thrombospondin motifs 13 (ADAMTS-13), and ADAMTS-like protein 1 (ADAMTSL-1) [reviewed in Heinonen and Maki 2009]. Thrombospondin is involved in the activation of latent TGFβ and the promoter of *B3GLCT* suggest activation in response to TGFβ as it contains SMAD-binding sites (1.1.3.1, Figure 7) [Heinonen *et al*. 2003]. This

suggests a role in regulation of TGFβ signalling and consequently possibly also in AD development. In order to asses whether patient 30544 could have an additional mutation in *B3GLCT* leading to a compound heterozygous state, Sanger sequencing was performed for all 15 exons and flanking intronic regions in this patient, but no further mutation was detected. In our NGS data no homozygous possibly pathogenic sequence variants have been detected, suggesting no major role of this gene in our patient cohort (Table 10). However, further mutation screening for patient 30544 using our NGS panel revealed no promising sequence variant in our set of potential AD candidate genes indicating the involvement of novel AD genes and consequently the need for further research (Table 9).

Two deletions affect the protocadherin-gamma gene cluster (PCDHG). This cluster involves 22 tandemly arranged genes, which have all unique exon 1 and share exons 2-4 [Wu and Maniatis 1999]. A deletion of 10,243 bp in patient 127 encompasses parts of exon 1 of *PCDHGB4* and exon 1 of *PCDHGB5* as well as complete exon 1 of *PCDHGA8* leading to a fusion exon (Figure 14). Due to high homology of these two exons break and rejoining could have occurred at 402 different positions, of which Chr5:140,767,901-140,778,144 is the most telomeric location of the deletion. As the transcription of individual genes of this cluster is regulated by distinct promoters upstream of each variable exon [Wang *et al*. 2002], the fusion exon, which corresponds to the sequence of *PCDHGB5,* will be likely under the control of the promoter of *PCDHGB4*, which is, unlike *PCDHGB5*, not only expressed in brain but most frequently also in fibroblasts (www.gtexportal.org/home/gene/PCDHGB4) [Matsuyoshi and Imamura 1997]. Members of the cadherin superfamily, to which the protocadherins belong, mediate calcium-dependent cell adhesion and cell signalling as well as are involved in development and tissue morphogenesis [reviewed in Halbleib and Nelson 2006]. As protocadherins expressed outside the brain are not well studied yet, a function in connective tissue development and thus a role in AD development cannot be excluded (cf. www.gtexportal.org/home/gene/PCDHG). However, a similar deletion has also been reported in a healthy Asian individual [Park *et al*. 2010]. Furthermore, patient 127 has a number of novel missense variants in our AD candidate genes, which are predicted to be disease causing and have to be followed up in more detail (Table 9), and no novel sequence variants have been detected in *PCDHGB4* in our NGS data (Table 10). This all together suggests that this deletion has a rather low if any impact on the AD phenotype of patient 127.

The second deletion in this cluster was detected in patient 37, comprises 4,041 bp (Chr5:140,801,272-140,805,312), and affects the 3'-end of exon 1 of *PCDHGA11*. As this deletion was also detected in the healthy mother of the index patient, clinical consequences of this deletion are improbable, but a modifier effect cannot be excluded (data not shown; cf. www.gtexportal.org/home/gene/PCDHGA11). This conclusion is supported by duplications in the deleted region reported in the Database of Genomic Variants (DGV, http://dgv.tcag.ca).

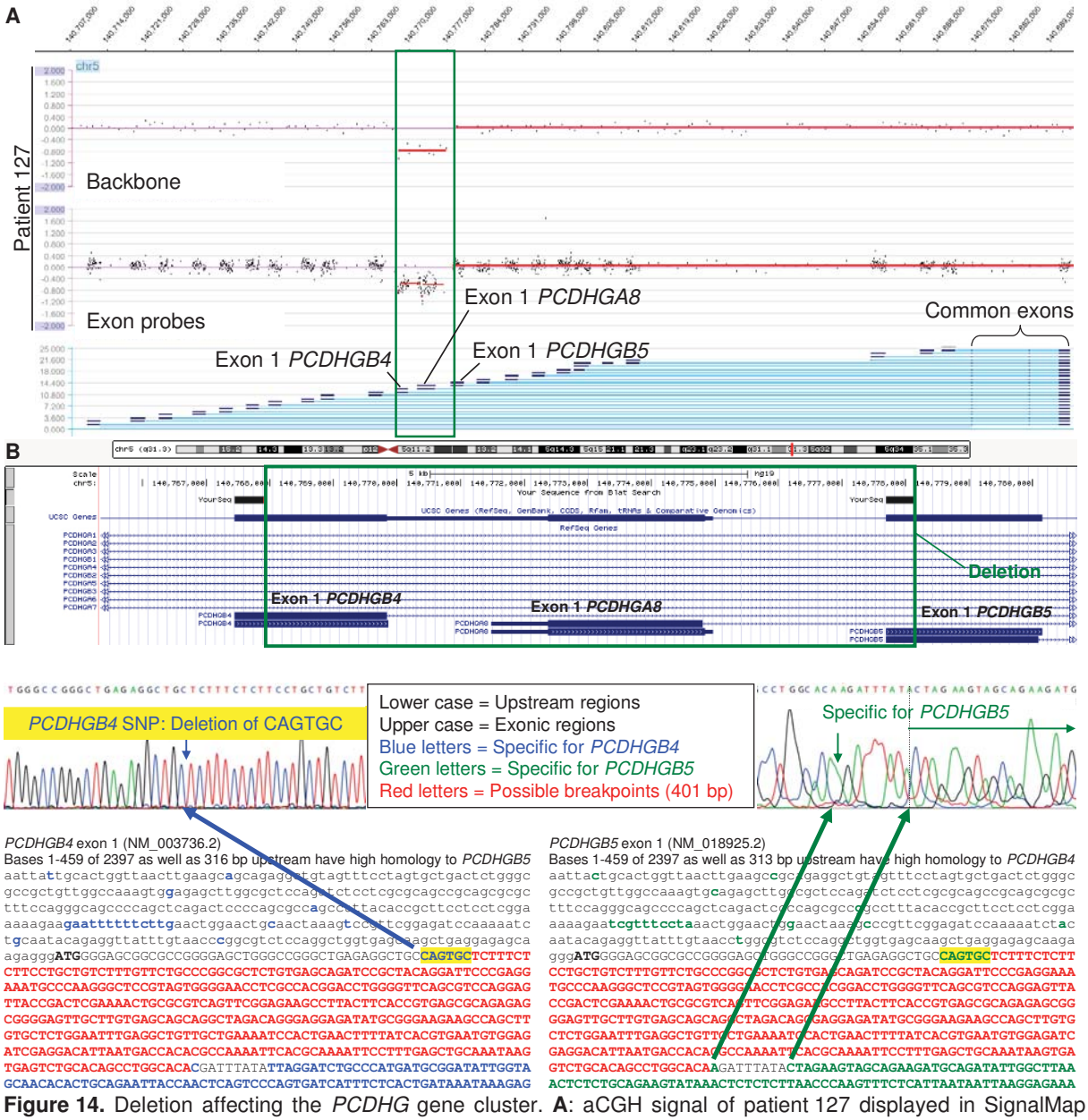



**Figure 14.** Deletion affecting the *PCDHG* gene cluster. **A**: aCGH signal of patient 127 displayed in SignalMap (Roche Diagnostics, Risch, Switzerland). **B**: Sanger sequences spanning the breakpoints and deleted region displayed in the UCSC Genome Browser (http://genome.ucsc.edu).

Patient 361 has a deletion of 1,145 bp (Chr16:22,156,877-22,158,021, most telomeric position of 46 possible) including complete exon 26 of the gene *VWA3A*, which leads to frameshift and a premature stop codon (Figure 15). NMD leading to functional haploinsufficiency is expected but not yet confirmed. This gene encodes the protein von Willebrand factor A domain containing 3A with still unknown function. As the von Willebrand A domain is known to be involved in cell adhesion, ECM proteins, and integrin receptors [Whittaker and Hynes 2002], this gene is a potential candidate gene for AD. Neither in our NGS data in this gene nor in patient 361 in further AD candidate genes novel missense mutations at conserved nucleotide positions predicted to be disease causing have been detected, leaving the impact of this gene on AD still open (Table 9, Table 10).

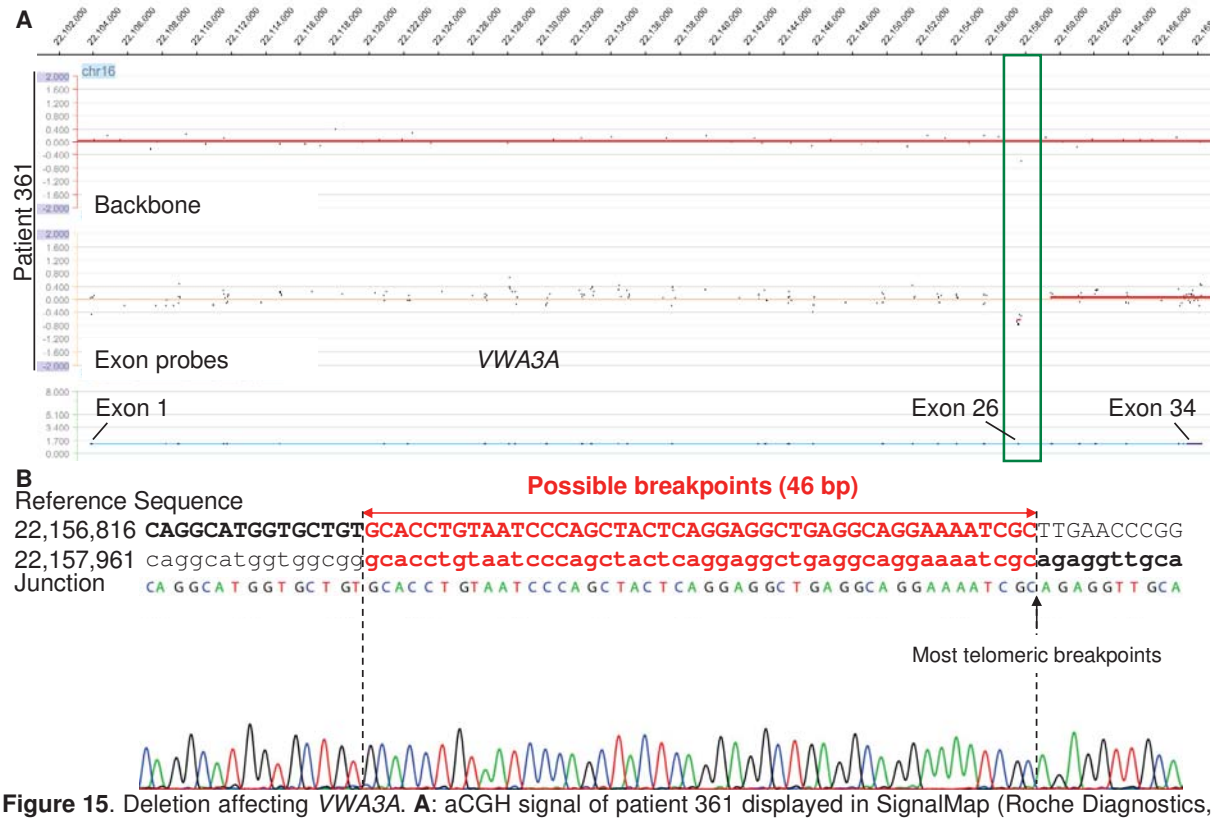



**Figure 15**. Deletion affecting *VWA3A*. **A**: aCGH signal of patient 361 displayed in SignalMap (Roche Diagnostics, Risch, Switzerland). **B**: Sanger sequences spanning the breakpoints.

In addition, promising deletions affecting genes encoding a metalloproteinase (*MMP26*), an integrin (*ITGAE*), a member of the mitochondrial respiratory chain (*NDUFA6*), and a myosin (*MYO19*) have been confirmed by LR-PCR, but the breakpoints were not covered using Sanger sequencing and PCR primers. Accordingly, these cases are perfect candidates for our developed procedure for breakpoint characterization [Okoniewski, Meienberg *et al*. 2013 (2.1.1)]. For two further deletions affecting genes encoding a cell adhesion molecule (*CNTNAP2*) and a collagen (*COL6A5*), which are potential AD candidate genes, confirmation is still ongoing as due to the size of the region of potential breakpoints the resulting fragments are too large for amplification by LR-PCR and additional analyses using internal primers or an different approach e.g. using mate-pair sequencing or WGS are needed (Table 8).

### 2.2.1.4 Discussion

Custom-designed and exome-focused aCGH was used to evaluate the role of middle sized and large deletions in our cohort of AD patients with unknown molecular basis. A first goal was to test the role of such deletions, which may be missed by standard genetic testing, in known AD genes. In our cohort of 65 patients no such deletion was detected leading to an expected frequency of 0/65 (0-5.6%, *P*=0.05). The second aim was to assess the role of our set of AD candidate genes, which was selected according to literature, meeting abstracts, and mouse models, as well as to detect novel candidate genes for AD. As deletions

comprising one or multiple exons can be considered as *a priori* pathogenic, genome-wide screening for such genetic aberrations is a useful approach for the identification of disease-causing genes as previously demonstrated for other genetic disorders [reviewed in Shinawi and Cheung 2008]. Accordingly, we expect for all detected deletions, which affect the coding region of a gene, an effect on its function. However, the remaining question is the impact of reduced function of this gene on clinical phenotype. It may be that one functional copy of the gene is sufficient, either because dose does not matter or higher expression of the remaining allele may compensate for the truncated copy. We detected in 11/65 patients (9.7-25.8%, $P$=0.05) a deletion in genes, with a potential function in connective tissue, contractile apparatus or TGFβ signalling, which might play a role in the pathogenesis of AD. Even though all these deletions overlap with larger known CNVs, which affect multiple genes and are reported in DGV, a pathogenic effect of these rather small deletions affecting in most of the cases only parts of the genes cannot be excluded. Accordingly, further analyses are needed for the assessment of the impact of these deletions and a potential role in the pathogenesis of AD for the affected genes. If not yet available, family members of the patients with the deletion or a promising sequence variant detected by NGS in this gene, have to be acquired to perform segregation analysis. Further options would be *in vitro* or even *in vivo* analyses.

Since the interpretation of the impact of such smaller deletions affecting just a few genes or even only parts of genes is challenging and work intensive, aCGH for diagnostic purposes is mainly applied in children with malformations and/or intellectual disability, where large SVs with the involvement of multiple genes are expected [reviewed in Kang and Koo 2012]. This is also the reason why in commercially available aCGH the resolution is only up to around 1 kb [Le Scouarnec and Gribble 2012]. Consequently, we used custom-designed arrays with a backbone of commercially available probes distributed across the whole genome and custom-designed probes to achieve higher resolution in exonic regions. This approach allowed us to detect also smaller deletions affecting only one or even only parts of an exon. We observed that the custom-designed probes are less stable than commercially available and are thus more prone to false positive calls (Figure 16A). Likewise, our design with both commercially available and custom-designed probes proved to be a useful approach as it combined the stability of the commercially available probes to reduce the number of false positives and the high resolution to detect also smaller deletions. However, also the source of DNA samples may influence the quality of aCGH results like in the case of DNA extracted from fibroblasts resulting in higher variance in exonic regions due to RNA contamination (Figure 16B), which can be solved by treating the DNA with RNase (Figure 16C).

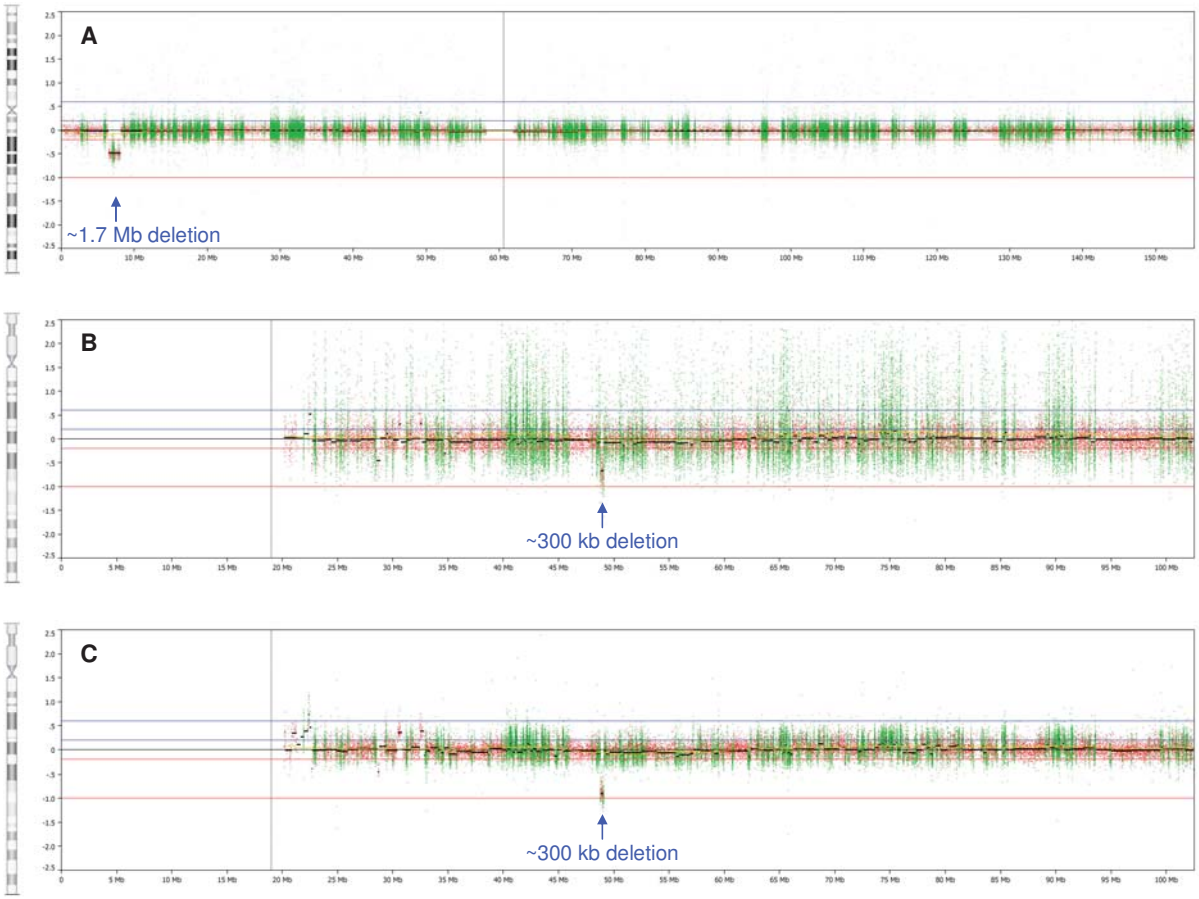



**Figure 16**. Effect of probe design and DNA source in aCGH. **A**: Chromosome X in a DNA sample extracted from EDTA-anticoagulated blood. **B**/**C**: Chromosome 15 in a DNA sample extracted from fibroblasts before (**B**) and after (**C**) RNase treatment. aCGH data are displayed using the software Nexus Copy Number 7 (Biodiscovery, Hawthorne, CA, USA) with red dots indicating commercially available backbone probes and green dots custom-designed probes to enrich exonic regions. Note that green dots have a larger range of variation than red dots, indicating lower stability.

Interestingly, all characterized deletions, except for the complex one in *B3GLCT*, contain stretches of 5-402 bp identical sequences at both ends of the deletions (Figure 12, Figure 13, Figure 14, and Figure 15), a phenomenon that has also been reported for other deletions [Giacalone and Francke 1992, Matyas *et al*. 2007, Meienberg *et al*. 2010 (Appendix 1)]. These identical sequences may have favoured the development of the deletion by one of the common mechanisms responsible for CNVs in the human genome, such as non-allelic homologous recombination (NAHR), non-homologous end joining (NHEJ), and Fork Stalling and Template Switching (FoSTeS) [Gu *et al*. 2008].

We also detected a number of novel deletions in non-coding or intergenic regions as well as further deletions in genes with unknown functions. Since also such regions can have an influence on gene expression [Qu and Fang 2013] and there is also evidence that CNVs can have intra- and inter-chromosomal effects on other genes due to interactions between chromosomal regions [Henrichsen *et al*. 2009, Lupianez *et al*. 2015], these deletions could also have an effect on the phenotype of our patients and need to be analysed in further detail, especially the ones close to a known AD gene or a potential candidate gene for AD. For the patients without a promising deletion, the disease-causing gene defect is expected to

be a SNV or an INDEL in a not yet sequenced gene. Furthermore, copy neutral variants such as inversions or translocations, which could also disrupt genes and consequently affect their function, are not detectable by aCGH and can thus not be excluded in these patients. Since these by aCGH not analysed sequence variant types are covered by WGS, which allows not only the identification of SNVs and INDELs, but also the detection and characterization of all kind of SVs on base pair level [Royer-Bertrand and Rivolta 2015], WGS will be the method of choice for further research to elucidate the underlying genetic defect in our cohort with unsolved AD cases.

# 3 General Discussion

## 3.1 Methodological Aspects

During this thesis, different state-of-the-art approaches have been evaluated and applied in patients suspected to be affected by a hereditary form of AD. One of the aims was to identify the underlying genetic defect of AD cases with unknown aetiology, thereby extending the knowledge on the molecular basis of aortic diseases.

For this aim, the first approach was the screening for large deletions using a custom-designed aCGH with probes enriched for exonic regions. This served as a whole-exome MLPA assay, enabling the genome-wide detection of one-exon deletions (2.2.1). Screening for large deletions can be a useful approach to identify new AD candidate genes as deletions affecting one or multiple exons of one or several genes can *a priori* be expected to be pathogenic. The use of this approach was also demonstrated in an unrelated project, where we were able to solve the case of an 11-year-old boy with muscle hypotonia, ataxia, therapy-resistant epilepsy, developmental delay, mental retardation, severe kyphoscoliosis, right ventricular hypertrophy, and gluten hypersensitivity. Postnatal karyotyping was negative but, thanks to the higher resolution of aCGH, we were able to identify a *de novo* deletion of ~10 Mb affecting around 50 genes (Appendix 4).

Furthermore, our aCGH data will be used to optimize the CNV calling algorithms for NGS. There are many commercial and freely available tools for the detection of CNVs in NGS data [reviewed in Xi *et al*. 2012]. However, these tools need thorough evaluation and most of them also optimization. Many of them are not so simple to use and have quite long running times. This is the reason why we developed our own approach analogous to the analysis of MLPA data. Like aCGH, this approach allows the detection of deletions and insertions as a stretch of multiple consecutive exons with relative read depth below and above a threshold, respectively [Meienberg, Zerjavic *et al*. 2015 (2.1.2)]. Using NGS data for the detection of CNVs would allow the detection of a broad spectrum of mutations in one assay. WGS is in this perspective much more powerful than WES or panel sequencing as it allows not only the detection of CNVs by changes in read depth, but also through consideration of insert size, split reads, and *de novo* assembly (Figure 11). Furthermore, read depth may be influenced by capture efficiency and PCR artefacts for GC-rich regions during library preparation, which is less an issue in PCR-free WGS. An additional advantage of WGS is the resolution of CNVs on base pair level, whereas in aCGH and targeted sequencing, including WES, additional methods for breakpoint characterization are needed [reviewed in Escaramis *et al*. 2015; cf. Okoniewski, Meienberg *et al*. 2013 (2.1.1)].

As the characterization of deletions using LR-PCR and Sanger sequencing may be laborious and time consuming, when internal primers are needed to cover the entire (LR-)PCR fragment, we evaluated the use of NGS to overcome this bottleneck [Okoniewski, Meienberg *et al*. 2013 (2.1.1)]. We were able to show that both second-generation sequencers with rather short reads and third-generation sequencers with long reads covering the complete (LR-)PCR fragment are useful for the characterization of large deletions. Likewise, for the characterization of deletions detected by our aCGH approach, where the breakpoints have not been detected (Table 8), such an NGS-based approach can be used.

A further approach for solving AD cases with unknown aetiology is the screening for SNVs and INDELs using NGS. Accordingly, we designed a protocol for NGS evaluation (Figure 17) and assessed the performance of different WES enrichment platforms, also in comparison with WGS, providing new insights into the performance of these powerful technologies [Meienberg, Zerjavic *et al*. 2015 (2.1.2)].

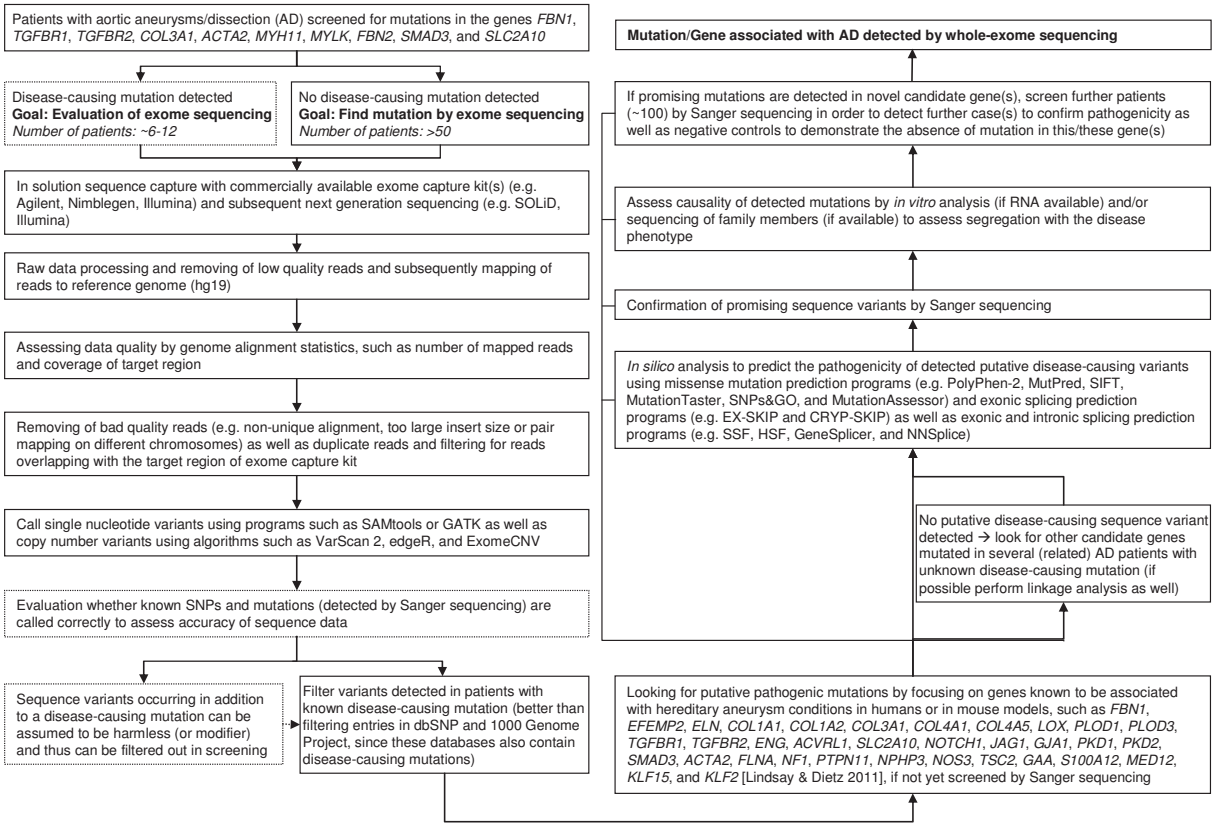
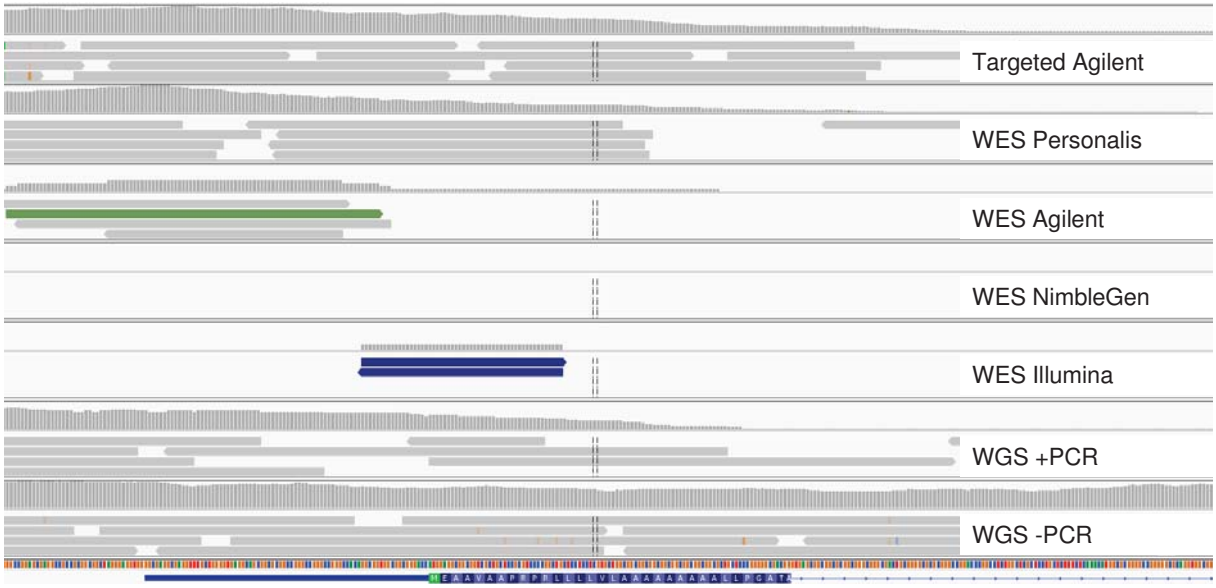



**Figure 17**. Overview of our study design for exome sequencing data evaluation and analysis (boxes with dotted lines represent evaluation steps). SAMtools, http://samtools.sourceforge.net [Li 2011]; GATK, Genome Analysis Toolkit (https://www.broadinstitute.org/gatk) [McKenna *et al*. 2010]; VarScan 2, http://varscan.sourceforge.net [Koboldt *et al*. 2012]; edgeR, http://www.bioconductor.org/packages/release/bioc/html/edgeR.html [Robinson *et al*. 2010]; ExomeCNV, https://secure.genome.ucla.edu/index.php/ExomeCNV_User_Guide [Sathirapongsasuti *et al*. 2011]; dbSNP, http://www.ncbi.nlm.nih.gov/SNP; 1000 Genome Project, http://www.1000genomes.org; PolyPhen-2, http://genetics.bwh.harvard.edu/pph2 [Adzhubei *et al*. 2010]; MutPred, http://mutpred.mutdb.org [Li *et al*. 2009]; SIFT, Sorting Intolerant From Tolerant (http://sift.jcvi.org) [Kumar *et al*. 2009]; MutationTaster, http://www.mutationtaster.org [Schwarz *et al*. 2014]; SNPs&GO, http://snps-and-go.biocomp.unibo.it/snps-and-go [Calabrese *et al*. 2009]; MutationAssessor, http://mutationassessor.org [Reva *et al*. 2011]; EX-SKIP, http://ex-skip.img.cas.cz [Raponi *et al*. 2011]; CRYP-SKIP, http://cryp-skip.img.cas.cz [Divina *et al*. 2009]; SSF, Splicing Sequences Finder (http://www.umd.be/searchSpliceSite.html); HSF, Human Splicing Finder (http://www.umd.be/HSF) [Desmet *et al*. 2009]; GeneSplicer, https://ccb.jhu.edu/software/genesplicer [Pertea *et al*. 2001]; NNSplice, Splice Site Prediction by Neural Network (http://www.fruitfly.org/seq_tools/splice.html) [Reese *et al*. 1997].

Evaluation of such enrichment platforms is important as their performance may be different to its specifications. On one hand, definitions of specifications like coverage of exome or probe design may differ between the platforms prohibiting the comparison of the kits. Furthermore, actual probe location and effectively achieved coverage may differ as the capture efficiency is dependent on different factors like GC content. Accordingly, methods including capture and/or PCR steps during library preparation achieve lower coverage for GC-rich regions compared to capture- and PCR-free WGS (Figure 18).




*TGFBR1* exon 1 (83% GC)

**Figure 18**. Coverage achieved by different sequencing applications on an Illumina sequencing platform in a GC-rich region (*TGFBR1* exon 1). Coverage tracks (grey vertical bars) and a few reads (grey horizontal bars/arrows) are displayed by the Integrative Genomics Viewer (IGV, https://www.broadinstitute.org/igv). Coloured bars in coverage track, called mismatches; coloured bars in reads, mismatched bases, purple vertical dashes, insertions; black horizontal lines, deletions; coloured reads, mapping positions of paired reads does not match; Targeted Agilent, custom-designed gene panel (SureSelect, Agilent); WES Personalis, clinically focused exome (ACEv2, Personalis); WES Agilent, SureSelect Human All Exon kit v5+UTR (Agilent); WES NimbleGen, SeqCap EZ Exome (v3) +UTR (NimbleGen, Roche); WES Illumina, Nextera Rapid Capture Expanded Exome (Illumina); WGS +PCR, TruSeq Nano DNA Sample Preparation Kit (not PCR-free, Illumina); WGS –PCR, TruSeq PCR-Free Sample Preparation Kit (Illumina).

However, not only the capture platform and library preparation influence NGS results but also the sequencing platform can have a high impact on data quality. Likewise, longer reads as well as paired-end sequencing improves the alignment in difficult regions. Like in traditional Sanger sequencing, NGS may have difficulties in homopolymeric or repetitive regions. This can be due to PCR steps during library preparation, but it also depends on the used sequencing method (1.2.2) [reviewed in Nguyen and Burnett 2014]. Likewise, miscalling in such regions is much more pronounced upon PCR-based target enrichment like the MASTR kits of Multiplicom compared to hybridization based enrichment like the Agilent SureSelect panel and the evaluated WES platforms (Figure 19). Concerning platform performance, the IonTorrent sequencing platform, where the numbers of identical nucleotides incorporated at once are estimated from signal intensities, has a much higher

error rate of INDEL calls than Illumina, where each incorporated nucleotide gives one signal, whereas the outdated SOLiD sequencing platform suffers from uneven coverage (Figure 20).

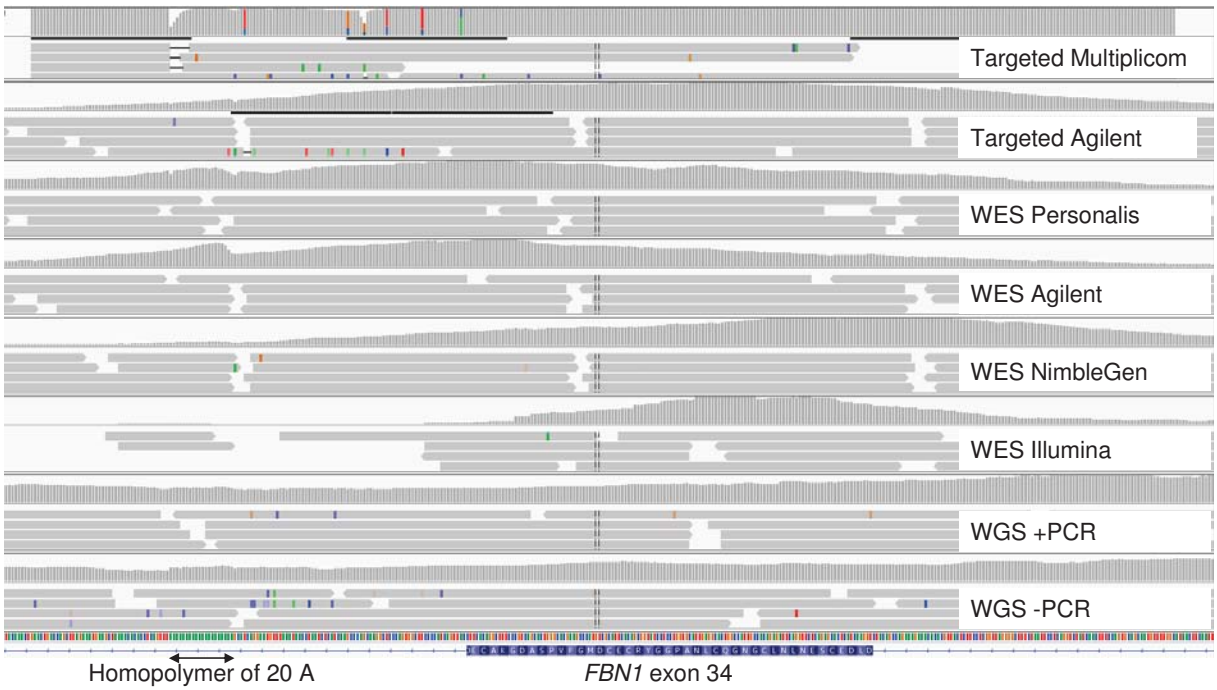



**Figure 19**. Reads produced for a region including a homopolymer of 20 bases (A) using different sequencing applications on an Illumina sequencing platform. Coverage tracks (grey vertical bars) and a few reads (grey horizontal bars/arrows) are displayed by the Integrative Genomics Viewer (IGV, cf. Figure 18). Targeted Multiplicom, MARFAN MASTR multiplex PCR kit (Multiplicom); Targeted Agilent, custom-designed gene panel (SureSelect, Agilent); WES Personalis, clinically focused exome (ACEv2, Personalis); WES Agilent, SureSelect Human All Exon kit v5+UTR (Agilent); WES NimbleGen, SeqCap EZ Exome (v3) +UTR (NimbleGen, Roche); WES Illumina, Nextera Rapid Capture Expanded Exome (Illumina); WGS +PCR, TruSeq Nano DNA Sample Preparation Kit (not PCR-free, Illumina); WGS –PCR, TruSeq PCR-Free Sample Preparation Kit (Illumina).

Further difficulties in NGS are genomic regions with high sequence homology as reads may have perfect matches at multiple positions in the genome. This is mainly an issue of short read length and is more pronounced in single-end compared to paired-end sequencing. Current sequence aligners have different strategies to deal with this issue such as discarding these reads, aligning just randomly to one of these positions or reporting multiple locations. Most of the gaps in WGS are due to this alignment difficulty [Treangen and Salzberg 2011]. One possibility to overcome this issue is to complement the short reads of the high throughput second-generation sequencing platforms with long reads generated with a different method like e.g. the third-generation sequencer PacBio, which achieves a read length of several kb [Chaisson *et al*. 2015]. In addition, aligners have to be optimized for such regions. There might be an option to create special alignment and variant calling settings for these regions, which allow reads to be mapped to multiple locations. It is of importance that such reads will be marked as reads with multiple alignments and will also be treated differently in variant calling as changed/reduced allele fractions have to be expected. This is mainly important when clinically relevant genes are affected, like in the case of *TNXB*, which is associated with Ehlers-Danlos syndrome hypermobility type (EDS III) [Zweers *et al*. 2003].

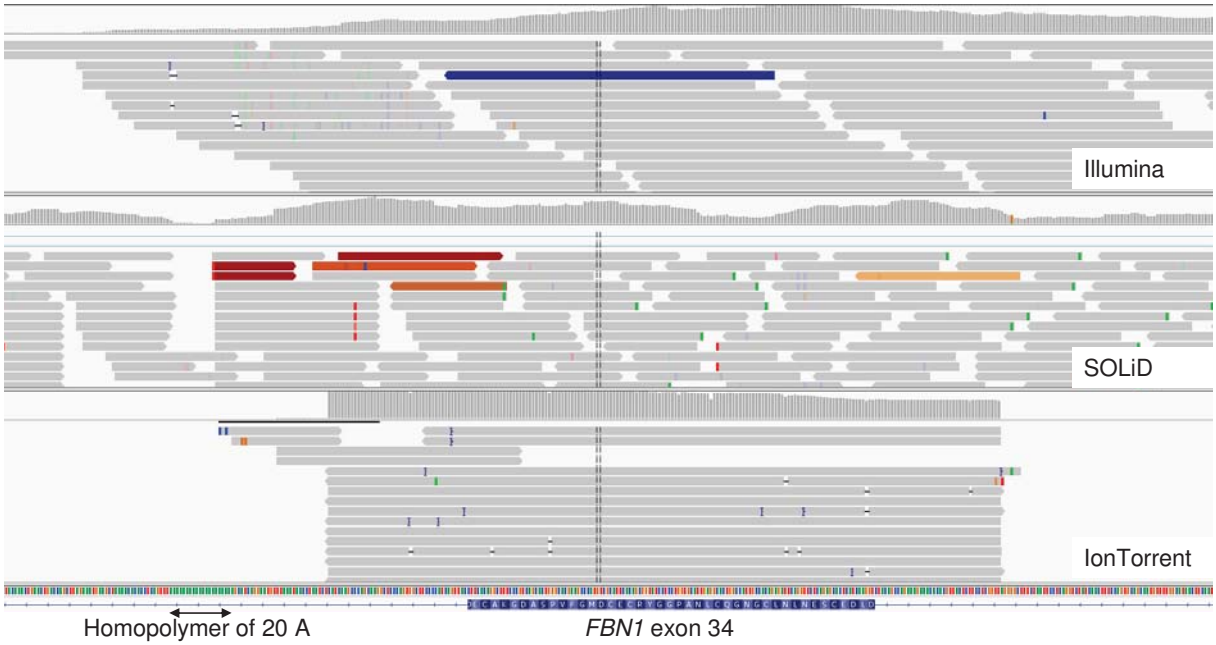



**Figure 20**. Reads produced for a region including a homopolymer of 20 bases (A) by WES using different sequencing platforms. Coverage tracks (grey vertical bars) and a few reads (grey horizontal bars/arrows) are displayed by the Integrative Genomics Viewer (IGV, cf. Figure 18). Illumina, SureSelect Human All Exon Kit v4+UTR (Agilent); SOLiD, SureSelect Human All Exon Kit v3 (Agilent); IonTorrent, IonAmpliSeq Exome Kit (Life Technologies, provided by Life Technologies in October 2013).

A wide range of different alignment and variant calling tools using different algorithms are available for the analysis of NGS data as freely or commercially available software solutions, resulting in slightly different alignments and variant calls. Each of it has its strength and limitations and none of it is perfect, making it a difficult task to determine the best analysis pipeline for a particular application. This issue was recognised by the Genome in a Bottle Consortium (http://genomeinabottle.org), which provides reference material as well as a Genome Comparison and Analytic Testing platform (GCAT, http://www.bioplanet.com/gcat) that allows to compare different alignment and variant calling pipelines [Zook *et al*. 2014, Highnam *et al*. 2015].

Once the alignment and variant calling is done, the next challenge is data interpretation, especially for WES and WGS. Likewise, many different tools are available for annotation, filtering, and prioritising detected/called sequence variants using different criteria like listing in mutation databases, phenotypic features of the patient, frequency of sequence variant in normal population, evolutionary conservation, and results of *in silico* prediction programs [Dolled-Filhart 2013]. One limitation of these software solutions is that they are only as accurate as the underlying databases/knowledge where the information is taken from. In addition to human databases, an additional approach would be to look at phenotypic databases with animal studies to detect in this way mutations in a relevant gene not yet associated with the human phenotype like used in the software exomiser [Robinson *et al*. 2014].

Further issues of WES and WGS are so called incidental findings. This means the detection of mutations associated with diseases (apparently) unrelated to the primary diagnostic question. These can be late-onset diseases like TAAD or Huntington disease, where symptoms develop with higher age, or predisposition to certain types of cancer, e.g. mutations in *BRCA1* and *BRCA2* genes which lead to an increased risk for breast and ovarian cancer. The American College of Medical Genetics and Genomics (ACMG, https://www.acmg.net) released recommendations how to deal with such sequence variants and published a list of genes associated with actionable phenotypes, i.e. where presymptomatic knowledge of the disease can prevent fatal outcome [Green *et al*. 2013]. As they recommend that patients have to agree to the reporting of incidental findings to receive genome-wide genetic analysis as well as that the geneticists should actively look for mutations in the recommended genes and to report them, these guidelines are rather controversial [Allyse and Michie 2013].

An alternative approach to WES are focused exomes like the ACE clinical exome enrichment platform (Personalis), which includes probes for ~7600 clinically relevant genes or gene panels like our custom-designed set of AD genes (2.2.1.2, Appendix 3). Advantages of these approaches are reduced data amount, less challenging interpretation with lower numbers of sequence variants with unknown significance, and in the case of panels also a lower risk for incidental findings and higher coverage. Furthermore, there is more space for probe design optimization allowing better coverage of targeted regions, especially also in difficult regions like such with high GC content (Figure 18). The down side of these approaches is the restriction to a certain set of genes, especially in the case of gene panels. In disorders with a high genetic and phenotypic heterogeneity and a considerable number of yet unknown genes expected to be involved, frequent updates of gene panels are required when mutations in additional genes are associated with the phenotype and a substantial number of mutation negative patients has to be expected and hence re-sequenced.

In disorders associated with large genes and no mutation hotspot like it is the case for many of the connective tissue genes associated with a syndromic form of AD, sequencing of all exons and flanking intronic sequences using Sanger sequencing is laborious and time-consuming, especially in diagnostic labs not equipped for high throughput. In such cases enrichment kits for NGS of just one gene or a small number of genes associated with one particular disease may reduce the workload and enhance the sample throughput. Such a kit is offered for the *FBN1* gene, mutations in which cause MFS (Figure 6, Table 2), by Multiplicom and allows the amplification of all 65 exons in four multiplex PCRs. Such kits allow sequence analysis similar to Sanger sequencing. However, despite high coverage and thus higher confidence, PCR- and NGS-related issues like homopolymers remain and may lead to false positive variant calls (Figure 19).

## 3.2 Diagnostics and Treatment Possibilities for Aortic Diseases

Knowledge of the molecular basis of AD allows the molecular testing of family members which are asymptomatic or have unspecific symptoms, i.e. do not fulfil the diagnostic criteria. It is not only a huge relief for the concerning person to know what to face, but it will also help to reduce health costs as only affected family members need regular cardiovascular control examinations. A further aspect is that the molecular basis gives some information on the progression of disease and associated risks. Likewise, in EDS IV dissections and ruptures are not limited to the aorta, but can also occur in large and middle-sized arteries and are not always associated with aneurysms. Thus, in this case regular control is not able to prevent all emergency situations by timely recognition and adaptation of life style is even more important. In LDS, progression of aortic diameters can be faster than in MFS and the aorta can rupture on a smaller diameter. The knowledge of the aetiology of the disease is also important to get access to targeted therapies, which are only available for specific syndromes, like in the cases of losartan, which is so far only available for MFS patients [Attenhofer Jost *et al*. 2014]. Similarly, for EDS IV a recent clinical trial has demonstrated the use of the β-blocker celiprolol [Ong *et al*. 2010].

TGFβ signalling and MFS is also a good example how better insights into the signalling pathways involved in pathogenesis may lead to the development of a targeted therapy and the identification of novel genes associated with a related phenotype (1.1.3.1, 1.1.4). The knowledge on the involvement of TGFβ signalling in MFS led to the development of promising therapeutic strategies targeting this pathway like the use of the ARB losartan. However, not all MFS patients respond to losartan, which could be due to the nature of the mutation itself or due to modifying genetic factors. Likewise, a recent study showed that losartan has lower effect on blood pressure reduction but higher efficiency to reduce the growth rate of aortic diameter in patients with a *FBN1* mutation leading to haploinsufficiency compared to such with a mutation leading to a dominant negative effect [Franken *et al*. 2015]. Consequently, knowledge of the disease-causing sequence variant will help to stratify, which MFS patients should be on a losartan treatment that can also have an impact on life quality due to its side effects and which patients will benefit more from an alternative or combined therapy with β-blockers. Furthermore, the involvement of TGFβ signalling in the pathogenesis of MFS also led to the finding that mutations in genes encoding players of the TGFβ signalling pathway like the ligands (*TGFB2* and *TGFB3*), the receptors (*TGFBR1* and *TGFBR2*), and some mediators of the canonical signalling (*SMAD3*) are associated with LDS, which shows high phenotypic overlap with MFS (1.1.3.1, Figure 6, Figure 7). Surprisingly, mutations in these active players of TGFβ signalling are associated with enhanced TGFβ signalling rather than reduced as intuitively expected. The reason for this paradox is still a matter of research. It is hypothesized that this increased signalling comes

from a shift in relative amount of ligands with increased availability and signalling through TGFβ1, as in *TGFB1,* the gene encoding this ligand, so far no mutation has been associated with an MFS/LDS-related phenotype. Another possible explanation might be a shift in the signalling through the two different TGFβ signalling pathways (canonical and non-canonical). However, this is rather unlikely as it was shown that signalling trough both pathways may be increased in MFS [Gallo *et al*. 2014].

Knowledge of the molecular basis of a disease is also the first step to find a targeted therapy. Likewise, we described the first case of true haploinsufficiency for *COL3A1* leading to an atypical form of EDS IV with incomplete penetrance for AD [Meienberg *et al*. 2010 (Appendix 1)], for which we started a study to find targeted therapy (Figure 21, Appendix 5). Briefly, we will use a recently described mouse model with true haploinsufficiency for *Col3a1* due to a spontaneous deletion, which leads to reduced mechanical stability of the aorta and in ~28% of heterozygous mice to spontaneous rupture of the aorta and thus to increased mortality similar to the phenotype in human patients [Smith *et al*. 2011]. This mouse model will be treated with different candidate substances with the goal to increase the mechanical stability of the aorta and thereby reduce mortality. The mechanical stability of the aorta will be tested using the approach defined in our pilot experiments, where we were able to show that thoracic aortas from heterozygous mice rupture at a significantly lower force than the ones of wild-type mice and that in both genotypes the maximal force at rupture decreases with increasing distance from the heart (Appendix 6).

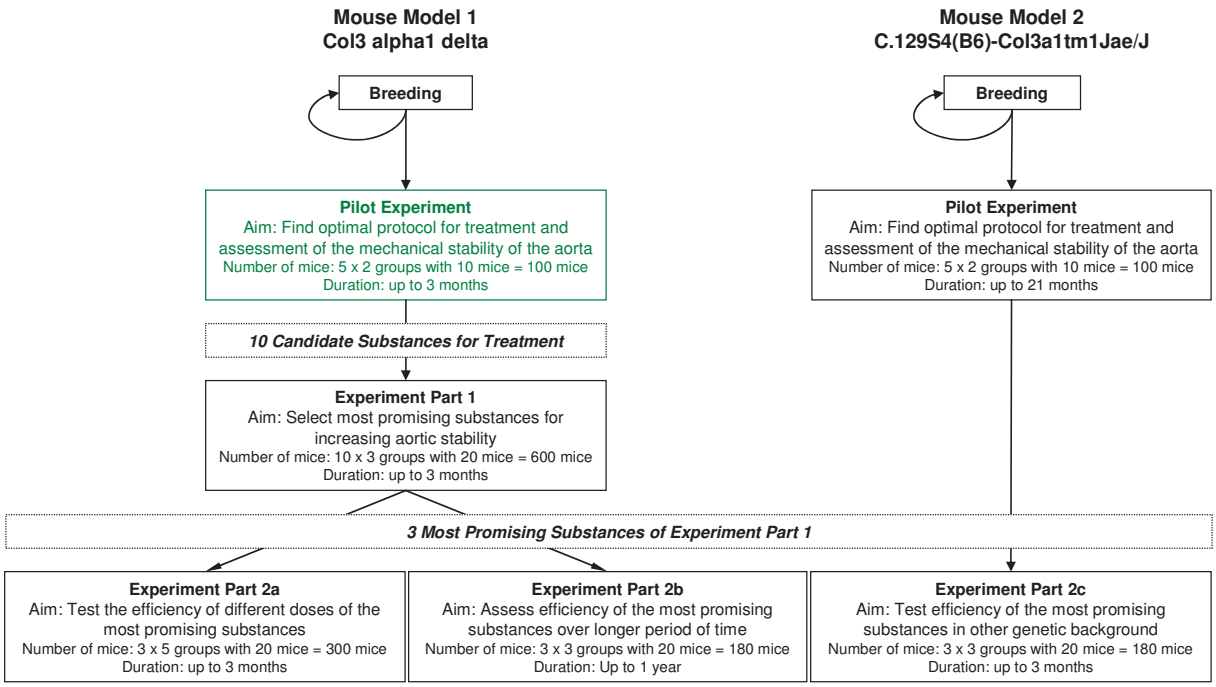



**Figure 21**. Overview of the planned mouse experiments. Green colour indicates completed parts.

A recent study showed that TGFβ signalling is not only elevated in MFS and related syndromes caused by mutations in genes involved in this pathway but also in EDS IV due to inflammation [Morissette *et al*. 2014]. Thus one of the tested substances will be the ARB losartan, which has not only the capability to reduce blood pressure but also to reduce TGFβ signalling and has been shown to ameliorate or even reverse the aortic phenotype in a mouse model of MFS [Habashi *et al*. 2006]. However, losartan is not only for its potential beneficial effect an essential substance to be tested in our mouse model but also due to potential adverse effects. Since increased TGFβ signalling also upregulates the expression of collagen, one could speculate that in EDS IV losartan could also have a negative effect worsening the aortic phenotype by decreasing the expression and thus the amount of collagen in the aortic tissue (1.1.3.1, Figure 7). The knowledge of such adverse effects would be crucial to avoid any harm in human patients as EDS IV patients could also clinically be misdiagnosed with MFS which are increasingly treated with losartan. Such results would show once more the importance of genetic testing prior to treatment.

A second approach will be to increase the amount of collagen by reducing its degradation by MMPs. Accordingly, we will inhibit MMP activity by applying doxycycline, for which a beneficial effect on the aortic phenotype has already been demonstrated in a mouse model of MFS and the first mouse model of EDS IV (1.1.4) [Xiong *et al*. 2008, Briest *et al*. 2011, Tae *et al*. 2012]. Thereby, it will be of interest to see whether the published results from a different mouse model for EDS IV can be confirmed with the mouse model used in our study despite different gene defect, genetic background, and approach to measure the mechanical stability of the aorta.

A third drug to be tested will be the β-blocker celiprolol. Recently, a clinical trial with this antihypertensive drug has been completed, demonstrating that in patients with EDS IV it leads to a prevention or delay of dissections and ruptures of the aorta. However, it was not examined whether or not this substance has an effect on the mechanical stability or the structure of the aortic wall [Ong *et al*. 2010].

However, when analysing the results of this study, it has to be kept in mind that mice are not small humans [Seok *et al*. 2013]. Interestingly, mice can produce vitamin C, whereas humans like some other higher mammals lost this capability [Chatterjee 1973]. Vitamin C is a cofactor of lysyl and prolyl hydroxylases and hence plays an important role in collagen synthesis [Szarka and Lorincz 2013]. The resulting difference in the availability of vitamin C between mice and human could consequently have an influence on collagen synthesis and hence also on the effect of tested drugs. A further aspect to consider is that mice studies are conducted in inbred strains. These inbred strains differ from each other at different polymorphic positions. Such SNPs could have a modifying effect on the studied phenotype or the effect of the tested substances [e.g. Sims *et al*. 2013]. We accounted for this by

backcrossing our mouse model which had a mixed background into the two corresponding inbred strains. In addition, we will also confirm the findings with the second available mouse model for EDS IV. If we see differences between the strains the big question will be, SNPs in which genes contribute to this difference. This knowledge will help to learn more about the disease mechanisms and modifying factors as well as may also be transmitted to humans.

Alternative therapeutic approaches based on the knowledge of the underlying disease or type of mutation may be gene therapies such as manipulating/adding long non-coding RNAs (lncRNAs). Such RNAs have recently been described and play an important role in the regulation of cell differentiation and in cardiovascular diseases [Uchida and Dimmeler 2015]. Another possibility would be to directly correct splicing defects or to suppress NMD in patients with mutations leading to functional haploinsufficiency by gene therapy [Bidou *et al*. 2012, Arechavala-Gomeza *et al*. 2014]. Recently, promising novel tools for RNA-guided site-specific induction of double-stranded breaks in target DNA, termed clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) systems, have been reported. These CRISPR/Cas systems enable facile, robust, and multiplexable systematic reverse engineering of causal genetic variants, a prerequisite for the development of targeted gene therapies [Jinek *et al*. 2012, Cong *et al*. 2013, Mali *et al*. 2013].

## 3.3 Outlook

It is still a long way to go to solve the bioinformatics limitations in the analysis and interpretation of NGS data. Moreover, the understanding of the genome is still incomplete and many open questions remain. Likewise, of the ~23,000 known protein-coding genes only ~15,000 are listed in the database Online Mendelian Inheritance in Men (OMIM, http://www.omim.org/statistics/entry, updated June 16, 2015) and even less are included in clinical exome sequencing kits (~7600 genes in ACEv2, 2.2.1.2), indicating the high number of genes with still unknown function. In addition, most of the mutations so far classified as disease-causing are located in the coding region or at conserved splice sites. With projects like ENCODE (ENCyclopedia Of DNA Elements) [ENCODE Project Consortium 2012] and the increased use of WGS, the knowledge on the non-coding regions is expected to increase and also mutations in this part of the genome will become better interpretable. Accordingly, a recently published study demonstrated the functional impact of genome architecture and the pathogenicity of SVs in non-coding regions disrupting boundaries of regulatory units of the human genome [Lupianez *et al*. 2015].

Despite these limitations is the revolution in molecular genetics with the implementation of NGS already ongoing and will change the general procedure in gene diagnostics. With the increased use of WES and WGS more cases with two or more weaker mutations leading to the phenotype will be detected. Likewise, the knowledge on modifying sequence variants will increase. With the novel sequencing technologies more cases with AD should get diagnosed and the time until diagnosis, which at the moment could be up to several years, should decrease. This will allow a more individualized follow-up and treatment of the patients. By identifying new genes associated with AD, the knowledge on pathogenesis and involved pathways will increase opening the way to the development of more targeted therapies.

# 4 References


Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR (2010) A method and server for predicting damaging missense mutations. Nat Methods 7:248-9

Allyse M, Michie M (2013) Not-so-incidental findings: the ACMG recommendations on the reporting of incidental findings in clinical whole genome and whole exome sequencing. Trends Biotechnol 31:439-41

Arechavala-Gomeza V, Khoo B, Aartsma-Rus A (2014) Splicing modulation therapy in the treatment of genetic diseases. Appl Clin Genet 7:245-52

Attenhofer Jost CH, Greutmann M, Connolly HM, Weber R, Rohrbach M, Oxenius A, Kretschmar O, Luscher TF, Matyas G (2014) Medical treatment of aortic aneurysms in Marfan syndrome and other heritable conditions. Curr Cardiol Rev 10:161-71

Barbier M, Gross MS, Aubart M, Hanna N, Kessler K, Guo DC, Tosolini L, Ho-Tin-Noe B, Regalado E, Varret M, Abifadel M, Milleron O, Odent S, Dupuis-Girod S, Faivre L, Edouard T, Dulac Y, Busa T, Gouya L, Milewicz DM, Jondeau G, Boileau C (2014) MFAP5 loss-of-function mutations underscore the involvement of matrix alteration in the pathogenesis of familial thoracic aortic aneurysms and dissections. Am J Hum Genet 95:736-43

Bertoli-Avella AM, Gillis E, Morisaki H, Verhagen JM, de Graaf BM, van de Beek G, Gallo E, et al. (2015) Mutations in a TGF-beta Ligand, TGFB3, Cause Syndromic Aortic Aneurysms and Dissections. J Am Coll Cardiol 65:1324-36

Bidou L, Allamand V, Rousset JP, Namy O (2012) Sense from nonsense: therapies for premature stop codon diseases. Trends Mol Med 18:679-88

Boileau C, Guo DC, Hanna N, Regalado ES, Detaint D, Gong L, Varret M, et al. (2012) TGFB2 mutations cause familial thoracic aortic aneurysms and dissections associated with mild systemic features of Marfan syndrome. Nat Genet 44:916-21

Briest W, Cooper TK, Tae HJ, Krawczyk M, McDonnell NB, Talan MI (2011) Doxycycline ameliorates the susceptibility to aortic lesions in a mouse model for the vascular type of Ehlers-Danlos syndrome. J Pharmacol Exp Ther 337:621-7

Brooke BS, Habashi JP, Judge DP, Patel N, Loeys B, Dietz HC, 3rd (2008) Angiotensin II blockade and aortic-root dilation in Marfan's syndrome. N Engl J Med 358:2787-95

Brunner NW, Ignaszewski A (2011) Aortic interlude: Dr Michael DeBakey, aortic dissection, and screening recommendations for abdominal aortic aneurysm. BCMJ 53:79-85

Buermans HP, den Dunnen JT (2014) Next generation sequencing technology: Advances and applications. Biochim Biophys Acta 1842:1932-1941

Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio R (2009) Functional annotations improve the predictive score of human disease-related mutations in proteins. Hum Mutat 30:1237-44

Chaisson MJ, Huddleston J, Dennis MY, Sudmant PH, Malig M, Hormozdiari F, Antonacci F, Surti U, Sandstrom R, Boitano M, Landolin JM, Stamatoyannopoulos JA, Hunkapiller MW, Korlach J, Eichler EE (2015) Resolving the complexity of the human genome using single-molecule sequencing. Nature 517:608-11

Chatterjee IB (1973) Evolution and the biosynthesis of ascorbic acid. Science 182:1271-2

Chiu HH, Wu MH, Wang JK, Lu CW, Chiu SN, Chen CA, Lin MT, Hu FC (2013) Losartan added to beta-blockade therapy for aortic root dilation in Marfan syndrome: a randomized, open-label pilot study. Mayo Clin Proc 88:271-6

Chrystoja CC, Diamandis EP (2014) Whole genome sequencing as a diagnostic test: challenges and opportunities. Clin Chem 60:724-33

Coady MA, Davies RR, Roberts M, Goldstein LJ, Rogalski MJ, Rizzo JA, Hammond GL, Kopf GS, Elefteriades JA (1999) Familial patterns of thoracic aortic aneurysms. Arch Surg 134:361-7

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F (2013) Multiplex genome engineering using CRISPR/Cas systems. Science 339:819-23

Coucke PJ, Willaert A, Wessels MW, Callewaert B, Zoppi N, De Backer J, Fox JE, Mancini GM, Kambouris M, Gardella R, Facchetti F, Willems PJ, Forsyth R, Dietz HC, Barlati S, Colombi M, Loeys B, De Paepe A (2006) Mutations in the facilitative glucose transporter GLUT10 alter angiogenesis and cause arterial tortuosity syndrome. Nat Genet 38:452-7

Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C (2009) Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res 37:e67

Dietz HC, Cutting GR, Pyeritz RE, Maslen CL, Sakai LY, Corson GM, Puffenberger EG, Hamosh A, Nanthakumar EJ, Curristin SM, Stetten G, Meyers DA, Francomano CA (1991) Marfan syndrome caused by a recurrent *de novo* missense mutation in the fibrillin gene. Nature 352:337-9

Divina P, Kvitkovicova A, Buratti E, Vorechovsky I (2009) Ab initio prediction of mutation-induced cryptic splice-site activation and exon skipping. Eur J Hum Genet 17:759-65

Dolled-Filhart MP, Lee M, Jr., Ou-Yang CW, Haraksingh RR, Lin JC (2013) Computational and bioinformatics frameworks for next-generation whole exome and genome sequencing. ScientificWorldJournal 2013:730210

Doyle AJ, Doyle JJ, Bessling SL, Maragh S, Lindsay ME, Schepers D, Gillis E, Mortier G, Homfray T, Sauls K, Norris RA, Huso ND, Leahy D, Mohr DW, Caulfield MJ, Scott AF, Destree A, Hennekam RC, Arn PH, Curry CJ, Van Laer L, McCallion AS, Loeys BL, Dietz HC (2012a) Mutations in the TGF-beta repressor SKI cause Shprintzen-Goldberg syndrome with aortic aneurysm. Nat Genet 44:1249-54

Doyle JJ, Gerber EE, Dietz HC (2012b) Matrix-dependent perturbation of TGFbeta signaling and disease. FEBS Lett 586:2003-15

ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489:57-74

Escaramis G, Docampo E, Rabionet R (2015) A decade of structural variants: description, history and methods to detect structural variation. Brief Funct Genomics

Franken R, den Hartog A, Radonic T, Micha D, Maugeri A, van Dijk FS, Meijers-Heijboer HE, Timmermans J, Scholte AJ, van den Berg MP, Groenink M, Mulder BJ, Zwinderman AH, de Waard V, Pals G (2015) Beneficial Outcome of Losartan Therapy Depends on Type of FBN1 Mutation in Marfan Syndrome. Circ Cardiovasc Genet

Gallo EM, Loch DC, Habashi JP, Calderon JF, Chen Y, Bedja D, van Erp C, Gerber EE, Parker SJ, Sauls K, Judge DP, Cooke SK, Lindsay ME, Rouf R, Myers L, ap Rhys CM, Kent KC, Norris RA, Huso DL, Dietz HC (2014) Angiotensin II-dependent TGF-beta signaling contributes to Loeys-Dietz syndrome vascular pathogenesis. J Clin Invest 124:448-60

Giacalone JP, Francke U (1992) Common sequence motifs at the rearrangement sites of a constitutional X/autosome translocation and associated deletion. Am J Hum Genet 50:725-41

Gibson MA, Finnis ML, Kumaratilake JS, Cleary EG (1998) Microfibril-associated glycoprotein-2 (MAGP-2) is specifically associated with fibrillin-containing microfibrils but exhibits more restricted patterns of tissue localization and developmental expression than its structural relative MAGP-1. J Histochem Cytochem 46:871-86

Gonzalez-Perez A, Lopez-Bigas N (2011) Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. Am J Hum Genet 88:440-9

Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, McGuire AL, Nussbaum RL, O'Daniel JM, Ormond KE, Rehm HL, Watson MS, Williams MS, Biesecker LG (2013) ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med 15:565-74

Groenink M, den Hartog AW, Franken R, Radonic T, de Waard V, Timmermans J, Scholte AJ, van den Berg MP, Spijkerboer AM, Marquering HA, Zwinderman AH, Mulder BJ (2013) Losartan reduces aortic dilatation rate in adults with Marfan syndrome: a randomized controlled trial. Eur Heart J 34:3491-500

Gu W, Zhang F, Lupski JR (2008) Mechanisms for human genomic rearrangements. Pathogenetics 1:4

Guo DC, Pannu H, Tran-Fadulu V, Papke CL, Yu RK, Avidan N, Bourgeois S, Estrera AL, Safi HJ, Sparks E, Amor D, Ades L, McConnell V, Willoughby CE, Abuelo D, Willing M, Lewis RA, Kim DH, Scherer S, Tung PP, Ahn C, Buja LM, Raman CS, Shete SS, Milewicz DM (2007) Mutations in smooth muscle alpha-actin (ACTA2) lead to thoracic aortic aneurysms and dissections. Nat Genet 39:1488-93

Guo DC, Regalado E, Casteel DE, Santos-Cortez RL, Gong L, Kim JJ, Dyack S, Horne SG, Chang G, Jondeau G, Boileau C, Coselli JS, Li Z, Leal SM, Shendure J, Rieder MJ, Bamshad MJ, Nickerson DA, Kim C, Milewicz DM (2013) Recurrent gain-of-function mutation in PRKG1 causes thoracic aortic aneurysms and acute aortic dissections. Am J Hum Genet 93:398-404

Guo DC, Gong L, Regalado ES, Santos-Cortez RL, Zhao R, Cai B, Veeraraghavan S, Prakash SK, Johnson RJ, Muilenburg A, Willing M, Jondeau G, Boileau C, Pannu H, Moran R, Debacker J, Bamshad MJ, Shendure J, Nickerson DA, Leal SM, Raman CS, Swindell EC, Milewicz DM (2015) MAT2A mutations predispose individuals to thoracic aortic aneurysms. Am J Hum Genet 96:170-7

Habashi JP, Judge DP, Holm TM, Cohn RD, Loeys BL, Cooper TK, Myers L, Klein EC, Liu G, Calvi C, Podowski M, Neptune ER, Halushka MK, Bedja D, Gabrielson K, Rifkin DB, Carta L, Ramirez F, Huso DL, Dietz HC (2006) Losartan, an AT1 antagonist, prevents aortic aneurysm in a mouse model of Marfan syndrome. Science 312:117-21

Habashi JP, Doyle JJ, Holm TM, Aziz H, Schoenhoff F, Bedja D, Chen Y, Modiri AN, Judge DP, Dietz HC (2011) Angiotensin II type 2 receptor signaling attenuates aortic aneurysm in mice through ERK antagonism. Science 332:361-5

Hadari YR, Gotoh N, Kouhara H, Lax I, Schlessinger J (2001) Critical role for the docking-protein FRS2 alpha in FGF receptor-mediated signal transduction pathways. Proc Natl Acad Sci U S A 98:8578-83

Halbleib JM, Nelson WJ (2006) Cadherins in development: cell adhesion, sorting, and tissue morphogenesis. Genes Dev 20:3199-214

Head SR, Komori HK, LaMere SA, Whisenant T, Van Nieuwerburgh F, Salomon DR, Ordoukhanian P (2014) Library construction for next-generation sequencing: overviews and challenges. Biotechniques 56:61-4, 66, 68, passim

Heinonen TY, Pasternack L, Lindfors K, Breton C, Gastinel LN, Maki M, Kainulainen H (2003) A novel human glycosyltransferase: primary structure and characterization of the gene and transcripts. Biochem Biophys Res Commun 309:166-74

Heinonen TY, Maki M (2009) Peters'-plus syndrome is a congenital disorder of glycosylation caused by a defect in the beta1,3-glucosyltransferase that modifies thrombospondin type 1 repeats. Ann Med 41:2-10

Henrichsen CN, Vinckenbosch N, Zollner S, Chaignat E, Pradervand S, Schutz F, Ruedi M, Kaessmann H, Reymond A (2009) Segmental copy number variation shapes tissue transcriptomes. Nat Genet 41:424-9

Highnam G, Wang JJ, Kusler D, Zook J, Vijayan V, Leibovich N, Mittelman D (2015) An analytical framework for optimizing variant discovery from personal genomes. Nat Commun 6:6275

Hodkinson BP, Grice EA (2015) Next-Generation Sequencing: A Review of Technologies and Tools for Wound Microbiome Research. Adv Wound Care (New Rochelle) 4:50-58

Holm TM, Habashi JP, Doyle JJ, Bedja D, Chen Y, van Erp C, Lindsay ME, Kim D, Schoenhoff F, Cohn RD, Loeys BL, Thomas CJ, Patnaik S, Marugan JJ, Judge DP, Dietz HC (2011) Noncanonical TGFbeta signaling contributes to aortic aneurysm progression in Marfan syndrome mice. Science 332:358-61

Humphrey JD, Milewicz DM, Tellides G, Schwartz MA (2014) Cell biology. Dysfunctional mechanosensing in aneurysms. Science 344:477-9

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337:816-21

Kang JU, Koo SH (2012) Evolving applications of microarray technology in postnatal diagnosis (review). Int J Mol Med 30:223-8

Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK (2012) VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 22:568-76

Kumar P, Henikoff S, Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4:1073-81

Lacro RV, Dietz HC, Sleeper LA, Yetman AT, Bradley TJ, Colan SD, Pearson GD, et al. (2014) Atenolol versus losartan in children and young adults with Marfan's syndrome. N Engl J Med 371:2061-71

Landenhed M, Engstrom G, Gottsater A, Caulfield MP, Hedblad B, Newton-Cheh C, Melander O, Smith JG (2015) Risk profiles for aortic dissection and ruptured or surgically treated aneurysms: a prospective cohort study. J Am Heart Assoc 4:e001513

Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, Maglott DR (2014) ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res 42:D980-5

Le Scouarnec S, Gribble SM (2012) Characterising chromosome rearrangements: recent technical advances in molecular cytogenetics. Heredity (Edinb) 108:75-85

Lesnik Oberstein SA, Kriek M, White SJ, Kalf ME, Szuhai K, den Dunnen JT, Breuning MH, Hennekam RC (2006) Peters Plus syndrome is caused by mutations in B3GALTL, a putative glycosyltransferase. Am J Hum Genet 79:562-6

Li B, Krishnan VG, Mort ME, Xin F, Kamati KK, Cooper DN, Mooney SD, Radivojac P (2009) Automated inference of molecular mechanisms of disease from amino acid substitutions. Bioinformatics 25:2744-50

Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27:2987-93

Lindsay ME, Dietz HC (2011) Lessons on the pathogenesis of aneurysm from heritable conditions. Nature 473:308-16

Lindsay ME, Schepers D, Bolar NA, Doyle JJ, Gallo E, Fert-Bober J, Kempers MJ, et al. (2012) Loss-of-function mutations in TGFB2 cause a syndromic presentation of thoracic aortic aneurysm. Nat Genet 44:922-7

Loeys BL, Chen J, Neptune ER, Judge DP, Podowski M, Holm T, Meyers J, Leitch CC, Katsanis N, Sharifi N, Xu FL, Myers LA, Spevak PJ, Cameron DE, De Backer J, Hellemans J, Chen Y, Davis EC, Webb CL, Kress W, Coucke P, Rifkin DB, De Paepe AM, Dietz HC (2005) A syndrome of altered cardiovascular, craniofacial, neurocognitive and skeletal development caused by mutations in TGFBR1 or TGFBR2. Nat Genet 37:275-81

Loeys BL, Schwarze U, Holm T, Callewaert BL, Thomas GH, Pannu H, De Backer JF, Oswald GL, Symoens S, Manouvrier S, Roberts AE, Faravelli F, Greco MA, Pyeritz RE, Milewicz DM, Coucke PJ, Cameron DE, Braverman AC, Byers PH, De Paepe AM, Dietz HC (2006) Aneurysm syndromes caused by mutations in the TGF-beta receptor. N Engl J Med 355:788-98

Loman NJ, Watson M (2015) Successful test launch for nanopore sequencing. Nat Methods 12:303-4

Lonigro RJ, Grasso CS, Robinson DR, Jing X, Wu YM, Cao X, Quist MJ, Tomlins SA, Pienta KJ, Chinnaiyan AM (2011) Detection of somatic copy number alterations in cancer using targeted exome capture sequencing. Neoplasia 13:1019-25

Lupianez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R, Santos-Simarro F, Gilbert-Dussardier B, Wittler L, Borschiwer M, Haas SA, Osterwalder M, Franke M, Timmermann B, Hecht J, Spielmann M, Visel A, Mundlos S (2015) Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell 161:1012-25

Majesky MW (2007) Developmental basis of vascular smooth muscle diversity. Arterioscler Thromb Vasc Biol 27:1248-58

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM (2013) RNA-guided human genome engineering via Cas9. Science 339:823-6

Matsuyoshi N, Imamura S (1997) Multiple cadherins are expressed in human fibroblasts. Biochem Biophys Res Commun 235:355-8

Matyas G, De Paepe A, Halliday D, Boileau C, Pals G, Steinmann B (2002) Evaluation and application of denaturing HPLC for mutation detection in Marfan syndrome: Identification of 20 novel mutations and two novel polymorphisms in the FBN1 gene. Hum Mutat 19:443-56

Matyas G, Arnold E, Carrel T, Baumgartner D, Boileau C, Berger W, Steinmann B (2006) Identification and in silico analyses of novel TGFBR1 and TGFBR2 mutations in Marfan syndrome-related disorders. Hum Mutat 27:760-9

Matyas G, Alonso S, Patrignani A, Marti M, Arnold E, Magyar I, Henggeler C, Carrel T, Steinmann B, Berger W (2007) Large genomic fibrillin-1 (FBN1) gene deletions provide evidence for true haploinsufficiency in Marfan syndrome. Hum Genet 122:23-32

Matyas G, Naef P, Tollens M, Oexle K (2014) De novo mutation of the latency-associated peptide domain of TGFB3 in a patient with overgrowth and Loeys-Dietz syndrome features. Am J Med Genet A 164A:2141-3

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297-303

McLoughlin D, McGuinness J, Byrne J, Terzo E, Huuskonen V, McAllister H, Black A, Kearney S, Kay E, Hill AD, Dietz HC, Redmond JM (2011) Pravastatin reduces Marfan aortic dilation. Circulation 124:S168-73

Meienberg J, Rohrbach M, Neuenschwander S, Spanaus K, Giunta C, Alonso S, Arnold E, Henggeler C, Regenass S, Patrignani A, Azzarello-Burri S, Steiner B, Nygren AO, Carrel T, Steinmann B, Matyas G (2010) Hemizygous deletion of COL3A1, COL5A2, and MSTN causes a complex phenotype with aortic dissection: a lesson for and from true haploinsufficiency. Eur J Hum Genet 18:1315-21

Meienberg J, Zerjavic K, Keller I, Okoniewski M, Patrignani A, Ludin K, Xu Z, Steinmann B, Carrel T, Rothlisberger B, Schlapbach R, Bruggmann R, Matyas G (2015) New insights into the performance of human whole-exome capture platforms. Nucleic Acids Res 43:e76

Milewicz DM, Guo DC, Tran-Fadulu V, Lafont AL, Papke CL, Inamoto S, Kwartler CS, Pannu H (2008) Genetic basis of thoracic aortic aneurysms and dissections: focus on smooth muscle cell contractile dysfunction. Annu Rev Genomics Hum Genet 9:283-302

Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, Abyzov A, et al. (2011) Mapping copy number variation by population-scale genome sequencing. Nature 470:59-65

Morissette R, Schoenhoff F, Xu Z, Shilane DA, Griswold BF, Chen W, Yang J, Zhu J, Fert-Bober J, Sloper L, Lehman J, Commins N, Van Eyk JE, McDonnell NB (2014) Transforming growth factor-beta and inflammation in vascular (type IV) Ehlers-Danlos syndrome. Circ Cardiovasc Genet 7:80-8

Mueller GC, Stierle L, Stark V, Steiner K, von Kodolitsch Y, Weil J, Mir TS (2014) Retrospective analysis of the effect of angiotensin II receptor blocker versus beta-blocker on aortic root growth in paediatric patients with Marfan syndrome. Heart 100:214-8

Neptune ER, Frischmeyer PA, Arking DE, Myers L, Bunton TE, Gayraud B, Ramirez F, Sakai LY, Dietz HC (2003) Dysregulation of TGF-beta activation contributes to pathogenesis in Marfan syndrome. Nat Genet 33:407-11

Newman WG, Black GC (2014) Delivery of a clinical genomics service. Genes (Basel) 5:1001-17

Nguyen L, Burnett L (2014) Automation of molecular-based analyses: a primer on massively parallel sequencing. Clin Biochem Rev 35:169-76

Nienaber CA, Clough RE (2015) Management of acute aortic dissection. Lancet 385:800-11

Okoniewski MJ, Meienberg J, Patrignani A, Szabelska A, Matyas G, Schlapbach R (2013) Precise breakpoint localization of large genomic deletions using PacBio and Illumina next-generation sequencers. Biotechniques 54:98-100

Oliver GR, Hart SN, Klee EW (2015) Bioinformatics for clinical next generation sequencing. Clin Chem 61:124-35

Ong KT, Perdu J, De Backer J, Bozec E, Collignon P, Emmerich J, Fauret AL, Fiessinger JN, Germain DP, Georgesco G, Hulot JS, De Paepe A, Plauchu H, Jeunemaitre X, Laurent S, Boutouyrie P (2010) Effect of celiprolol on prevention of cardiovascular events in vascular Ehlers-Danlos syndrome: a prospective randomised, open, blinded-endpoints trial. Lancet 376:1476-84

Pannu H, Fadulu VT, Chang J, Lafont A, Hasham SN, Sparks E, Giampietro PF, Zaleski C, Estrera AL, Safi HJ, Shete S, Willing MC, Raman CS, Milewicz DM (2005) Mutations in transforming growth factor-beta receptor type II cause familial thoracic aortic aneurysms and dissections. Circulation 112:513-20

Pardali E, Ten Dijke P (2012) TGFbeta signaling and cardiovascular diseases. Int J Biol Sci 8:195-213

Park H, Kim JI, Ju YS, Gokcumen O, Mills RE, Kim S, Lee S, et al. (2010) Discovery of common Asian copy number variants using integrated high-resolution array CGH and massively parallel DNA sequencing. Nat Genet 42:400-5

Pees C, Laccone F, Hagl M, Debrauwer V, Moser E, Michel-Behnke I (2013) Usefulness of losartan on the size of the ascending aorta in an unselected cohort of children, adolescents, and young adults with Marfan syndrome. Am J Cardiol 112:1477-83

Pertea M, Lin X, Salzberg SL (2001) GeneSplicer: a new computational method for splice site prediction. Nucleic Acids Res 29:1185-90

Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A (2010) Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res 20:110-21

Pope FM, Martin GR, Lichtenstein JR, Penttinen R, Gerson B, Rowe DW, McKusick VA (1975) Patients with Ehlers-Danlos syndrome type IV lack type III collagen. Proc Natl Acad Sci U S A 72:1314-6

Qu H, Fang X (2013) A brief review on the Human Encyclopedia of DNA Elements (ENCODE) project. Genomics Proteomics Bioinformatics 11:135-41
Quaglino D, Ronchetti IP (2002) The Cardiovascular System. In: Royce PM, Steinmann B (eds) Connective Tissue and its Heritable Disorders. Wiley-Liss, Inc, New York, pp 121-144
Ramensky V, Bork P, Sunyaev S (2002) Human non-synonymous SNPs: server and survey. Nucleic Acids Res 30:3894-900
Raponi M, Kralovicova J, Copson E, Divina P, Eccles D, Johnson P, Baralle D, Vorechovsky I (2011) Prediction of single-nucleotide substitutions that result in exon skipping: identification of a splicing silencer in BRCA1 exon 6. Hum Mutat 32:436-44
Reese MG, Eeckman FH, Kulp D, Haussler D (1997) Improved splice site detection in Genie. J Comput Biol 4:311-23
Reva B, Antipin Y, Sander C (2011) Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res 39:e118
Riegel M (2014) Human molecular cytogenetics: From cells to nucleotides. Genet Mol Biol 37:194-209
Rienhoff HY, Jr., Yeo CY, Morissette R, Khrebtukova I, Melnick J, Luo S, Leng N, Kim YJ, Schroth G, Westwick J, Vogel H, McDonnell N, Hall JG, Whitman M (2013) A mutation in TGFB3 associated with a syndrome of low muscle mass, growth retardation, distal arthrogryposis and clinical features overlapping with Marfan and Loeys-Dietz syndrome. Am J Med Genet A 161A:2040-6
Royer-Bertrand B, Rivolta C (2015) Whole genome sequencing as a means to assess pathogenic mutations in medical genetics and cancer. Cell Mol Life Sci 72:1463-71
Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139-40
Robinson PN, Kohler S, Oellrich A, Wang K, Mungall CJ, Lewis SE, Washington N, Bauer S, Seelow D, Krawitz P, Gilissen C, Haendel M, Smedley D (2014) Improved exome prioritization of disease genes through cross-species phenotype comparison. Genome Res 24:340-8
Saintenac C, Jiang D, Akhunov ED (2011) Targeted analysis of nucleotide and copy number variation by exon capture in allotetraploid wheat genome. Genome Biol 12:R88
Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 74:5463-7
Sathirapongsasuti JF, Lee H, Horst BA, Brunner G, Cochran AJ, Binder S, Quackenbush J, Nelson SF (2011) Exome sequencing-based copy-number variation and loss of heterozygosity detection: ExomeCNV. Bioinformatics 27:2648-54
Schadt EE, Turner S, Kasarskis A (2010) A window into third-generation sequencing. Hum Mol Genet 19:R227-40
Schwarz JM, Cooper DN, Schuelke M, Seelow D (2014) MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods 11:361-2
Seok J, Warren HS, Cuenca AG, Mindrinos MN, Baker HV, Xu W, Richards DR, et al. (2013) Genomic responses in mouse models poorly mimic human inflammatory diseases. Proc Natl Acad Sci U S A 110:3507-12
Shen Y, Wu BL (2009) Microarray-based genomic DNA profiling technologies in clinical molecular diagnostics. Clin Chem 55:659-69
Shinawi M, Cheung SW (2008) The array CGH and its clinical applications. Drug Discov Today 13:760-70
Shores J, Berger KR, Murphy EA, Pyeritz RE (1994) Progression of aortic dilatation and the benefit of long-term beta-adrenergic blockade in Marfan's syndrome. N Engl J Med 330:1335-41
Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, Weinstock GM, Wilson RK, Gibbs RA, Kent WJ, Miller W, Haussler D (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res 15:1034-50
Sims EK, Hatanaka M, Morris DL, Tersey SA, Kono T, Chaudry ZZ, Day KH, Moss DR, Stull ND, Mirmira RG, Evans-Molina C (2013) Divergent compensatory responses to high-fat diet between C57BL6/J and C57BLKS/J inbred mouse strains. Am J Physiol Endocrinol Metab 305:E1495-511
Smith LB, Hadoke PW, Dyer E, Denvir MA, Brownstein D, Miller E, Nelson N, Wells S, Cheeseman M, Greenfield A (2011) Haploinsufficiency of the murine Col3a1 locus causes aortic dissection: a novel model of the vascular type of Ehlers-Danlos syndrome. Cardiovasc Res 90:182-90
Steinmann B, Royce PM, Superti-Furga A (2002) The Ehlers-Danlos syndrome. In: Royce PM, Steinmann B (eds) Connective Tissue and its Heritable Disorders. Wiley-Liss, Inc., New York, pp 431-523
Superti-Furga A, Gugler E, Gitzelmann R, Steinmann B (1988) Ehlers-Danlos syndrome type IV: a multi-exon deletion in one of the two COL3A1 alleles affecting structure, stability, and processing of type III procollagen. J Biol Chem 263:6226-32
Swee W, Dake MD (2008) Endovascular management of thoracic dissections. Circulation 117:1460-73
Szarka A, Lorincz T (2013) The role of ascorbate in protein folding. Protoplasma 251:489-97
Tae HJ, Marshall S, Zhang J, Wang M, Briest W, Talan MI (2012) Chronic treatment with a broad-spectrum metalloproteinase inhibitor, doxycycline, prevents the development of spontaneous aortic lesions in a mouse model of vascular Ehlers-Danlos syndrome. J Pharmacol Exp Ther 343:246-51
Treangen TJ, Salzberg SL (2011) Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet 13:36-46
Tromp G, Kuivaniemi H, Shikata H, Prockop DJ (1989) A single base mutation that substitutes serine for glycine 790 of the alpha 1 (III) chain of type III procollagen exposes an arginine and causes Ehlers-Danlos syndrome IV. J Biol Chem 264:1349-52
Uchida S, Dimmeler S (2015) Long noncoding RNAs in cardiovascular diseases. Circ Res 116:737-50
van de Laar IM, Oldenburg RA, Pals G, Roos-Hesselink JW, de Graaf BM, Verhagen JM, Hoedemaekers YM, et al. (2011) Mutations in SMAD3 cause a syndromic form of aortic aneurysms and dissections with early-onset osteoarthritis. Nat Genet 43:121-6
van de Laar IM, van der Linde D, Oei EH, Bos PK, Bessems JH, Bierma-Zeinstra SM, van Meer BL, et al. (2012) Phenotypic spectrum of the SMAD3-related aneurysms-osteoarthritis syndrome. J Med Genet 49:47-57
van Dijk EL, Jaszczyszyn Y, Thermes C (2014a) Library preparation methods for next-generation sequencing: tone down the bias. Exp Cell Res 322:12-20
van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C (2014b) Ten years of next-generation sequencing technology. Trends Genet 30:418-26
Voelkerding KV, Dames SA, Durtschi JD (2009) Next-generation sequencing: from basic research to diagnostics. Clin Chem 55:641-58
Wang X, Su H, Bradley A (2002) Molecular mechanisms governing Pcdh-gamma gene expression: evidence for a multiple promoter and cis-alternative splicing model. Genes Dev 16:1890-905

Wang L, Guo DC, Cao J, Gong L, Kamm KE, Regalado E, Li L, Shete S, He WQ, Zhu MS, Offermanns S, Gilchrist D, Elefteriades J, Stull JT, Milewicz DM (2010) Mutations in myosin light chain kinase cause familial aortic dissections. Am J Hum Genet 87:701-7

Wilkie AO (2005) Bad bones, absent smell, selfish testes: the pleiotropic consequences of human FGF receptor mutations. Cytokine Growth Factor Rev 16:187-203

Whittaker CA, Hynes RO (2002) Distribution and evolution of von Willebrand/integrin A domains: widely dispersed domains with roles in cell adhesion and elsewhere. Mol Biol Cell 13:3369-87

Wu Q, Maniatis T (1999) A striking organization of a large family of human neural cadherin-like cell adhesion genes. Cell 97:779-90

Xi R, Lee S, Park PJ (2012) A survey of copy-number variation detection tools based on high-throughput sequencing data. Curr Protoc Hum Genet Chapter 7:Unit7 19

Xiong W, Knispel RA, Dietz HC, Ramirez F, Baxter BT (2008) Doxycycline delays aneurysm rupture in a mouse model of Marfan syndrome. J Vasc Surg 47:166-72; discussion 172

Xuan J, Yu Y, Qing T, Guo L, Shi L (2013) Next-generation sequencing in the clinic: promises and challenges. Cancer Lett 340:284-95

Zhu L, Vranckx R, Khau Van Kien P, Lalande A, Boisset N, Mathieu F, Wegman M, Glancy L, Gasc JM, Brunotte F, Bruneval P, Wolf JE, Michel JB, Jeunemaitre X (2006) Mutations in myosin heavy chain 11 cause a syndrome associating thoracic aortic aneurysm/aortic dissection and patent ductus arteriosus. Nat Genet 38:343-9

Zook JM, Chapman B, Wang J, Mittelman D, Hofmann O, Hide W, Salit M (2014) Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls. Nat Biotechnol 32:246-51

Zweers MC, Bristow J, Steijlen PM, Dean WB, Hamel BC, Otero M, Kucharekova M, Boezeman JB, Schalkwijk J (2003) Haploinsufficiency of TNXB is associated with hypermobility type of Ehlers-Danlos syndrome. Am J Hum Genet 73:214-7

# 5 Appendix

# Appendix 1 Publication: Hemizygous Deletion of *COL3A1*, *COL5A2*, and *MSTN* Causes a Complex Phenotype with Aortic Dissection: a Lesson for and from True Haploinsufficiency

Meienberg J, Rohrbach M, Neuenschwander S, Spanaus K, Giunta C, Alonso S, Arnold E, Henggeler C, Regenass S, Patrignani A, Azzarello-Burri S, Steiner B, Nygren AOH, Carrel T, Steinmann B, Matyas G (2010) Hemizygous deletion of *COL3A1*, *COL5A2*, and *MSTN* causes a complex phenotype with aortic dissection: a lesson for and from true haploinsufficiency. Eur J Hum Genet 18:1315-1321.

Impact factor: 4.380 (2010)

European Journal of Human Genetics (2010) 18, 1315–1321

www.nature.com/ejhg

ARTICLE

# Hemizygous deletion of *COL3A1*, *COL5A2*, and *MSTN* causes a complex phenotype with aortic dissection: a lesson for and from true haploinsufficiency

Janine Meienberg[1], Marianne Rohrbach[2], Stefan Neuenschwander[3], Katharina Spanaus[4], Cecilia Giunta[2], Sira Alonso[1], Eliane Arnold[1,2], Caroline Henggeler[1], Stephan Regenass[5], Andrea Patrignani[3], Silvia Azzarello-Burri[6], Bernhard Steiner[6], Anders OH Nygren[7], Thierry Carrel[8], Beat Steinmann[2] and Gábor Mátyás*[,1]

**Aortic dilatation/dissection (AD) can occur spontaneously or in association with genetic syndromes, such as Marfan syndrome (MFS; caused by *FBN1* mutations), MFS type 2 and Loeys–Dietz syndrome (associated with *TGFBR1*/*TGFBR2* mutations), and Ehlers–Danlos syndrome (EDS) vascular type (caused by *COL3A1* mutations). Although mutations in *FBN1* and *TGFBR1*/*TGFBR2* account for the majority of AD cases referred to us for molecular genetic testing, we have obtained negative results for these genes in a large cohort of AD patients, suggesting the involvement of additional genes or acquired factors. In this study we assessed the effect of *COL3A1* deletions/duplications in this cohort. Multiplex ligation-dependent probe amplification (MLPA) analysis of 100 unrelated patients identified one hemizygous deletion of the entire *COL3A1* gene. Subsequent microarray analyses and sequencing of breakpoints revealed the deletion size of 3 408 306 bp at 2q32.1q32.3. This deletion affects not only *COL3A1* but also 21 other known genes (*GULP1*, *DIRC1*, *COL5A2*, *WDR75*, *SLC40A1*, *ASNSD1*, *ANKAR*, *OSGEPL1*, *ORMDL1*, *LOC100129592*, *PMS1*, *MSTN*, *C2orf88*, *HIBCH*, *INPP1*, *MFSD6*, *TMEM194B*, *NAB1*, *GLS*, *STAT1*, and *STAT4*), mutations in three of which (*COL5A2*, *SLC40A1*, and *MSTN*) have also been associated with an autosomal dominant disorder (EDS classical type, hemochromatosis type 4, and muscle hypertrophy). Physical and laboratory examinations revealed that true haploinsufficiency of *COL3A1*, *COL5A2*, and *MSTN*, but not that of *SLC40A1*, leads to a clinical phenotype. Our data not only emphasize the impact/role of *COL3A1* in AD patients but also extend the molecular etiology of several disorders by providing hitherto unreported evidence for true haploinsufficiency of the underlying gene.**
*European Journal of Human Genetics* (2010) **18,** 1315–1321; doi:10.1038/ejhg.2010.105; published online 21 July 2010

**Keywords:** aorta; cardiovascular genetics; collagen; true haploinsufficiency

## INTRODUCTION

Aortic dilatation/dissection (AD) is a life-threatening condition associated with considerable morbidity and mortality. It can occur spontaneously, because of cardiovascular risk factors (eg, hypertension), or in association with genetic disorders, such as familial thoracic aortic aneurysms leading to type A dissections (TAAD, MIM #132900), Marfan syndrome (MFS, MIM #154700), Loeys–Dietz syndrome (LDS, MIM #609192), and the vascular type of Ehlers–Danlos syndrome (EDS IV, MIM #130050). MFS is an autosomal dominant systemic disorder of connective tissue (1–2:10 000).[1] It shows variable manifestations in the cardiovascular, skeletal, and ocular systems.[2] In the majority of cases, MFS is caused by mutations in the *FBN1* gene (MIM *134797), which encodes the extracellular matrix protein fibrillin-1.[3] The molecular etiology of MFS has been extended by the finding that heterozygous mutations in the genes encoding transforming growth factor-$\beta$ receptors I (*TGFBR1*, MIM *190181) and II (*TGFBR2*, MIM *190182) can lead to MFS-related disorders, such as MFS type 2 (MFS2; MIM #610380), LDS, and TAAD.[4–8]

EDS IV, the vascular type of EDS, is an autosomal dominant disorder of connective tissue as well, but less prevalent than MFS (~1–2:100 000).[9] In addition to complications in the cardiovascular system (arterial rupture), manifestations of EDS IV involve the skin, joints, and hollow organs, resulting in thin/translucent skin, extensive bruising, characteristic facial appearance, and intestinal/uterine rupture as major diagnostic criteria.[10] EDS IV is caused by mutations in *COL3A1* (MIM *120180), which encodes the $\alpha$1 chain of type III collagen.[11–14] So far, nearly 200 unique *COL3A1* mutations have been registered in locus-specific mutation databases (Human Gene Mutation Database, http://www.hgmd.cf.ac.uk; Leiden Open Variation Database, https://eds.gene.le.ac.uk), some of which led to functional haploinsufficiency by preferential degradation of mutant transcripts due to nonsense-mediated mRNA decay (NMD). However, no case of true *COL3A1* haploinsufficiency, that is, complete loss of

[1]Division of Medical Molecular Genetics and Gene Diagnostics, Institute of Medical Genetics, University of Zurich, Zurich, Switzerland; [2]Division of Metabolism, University Children's Hospital, Zurich, Switzerland; [3]Functional Genomics Center Zurich (FGCZ), ETH and University of Zurich, Zurich, Switzerland; [4]Institute for Clinical Chemistry, University Hospital, Zurich, Switzerland; [5]Division of Clinical Immunology, University Hospital, Zurich, Switzerland; [6]Division of Medical Genetics, Institute of Medical Genetics, University of Zurich, Zurich, Switzerland; [7]MRC Holland, Amsterdam, The Netherlands; [8]Clinic for Cardiovascular Surgery, University Hospital, Berne, Switzerland
*Correspondence: Dr G Mátyás, Division of Medical Molecular Genetics and Gene Diagnostics, Institute of Medical Genetics, University of Zurich, Schorenstrasse 16, CH-8603 Schwerzenbach, Switzerland. Tel: +41 44 6557031; Fax: +41 44 6557213; E-mail: matyas@medgen.uzh.ch
Received 11 January 2010; revised 6 May 2010; accepted 4 June 2010; published online 21 July 2010

one allele through hemizygous deletion, has been described at the molecular level before (cf. Supplementary Table S1 and Supplementary Figure S1).

In this study, we report a hemizygous deletion of *COL3A1* and flanking genes as well as assess the clinical and biochemical effects of this deletion. In addition to the importance of *COL3A1* mutations in AD patients, our results show the different role of true haploinsufficiency in the etiology of dominant disorders.

## MATERIALS AND METHODS

### Patients

A total of 100 unrelated AD patients with familial (~20/100) or sporadic (~80/100) phenotypes suggestive for TAAD/MFS/LDS/EDS IV and thus in some cases also involving the skeletal (~40/100) and/or ocular (~5/100) system were selected for this study. In this cohort, previous sequencing and multiplex ligation-dependent probe amplification (MLPA) analyses of *FBN1*, *TGFBR1*, and *TGFBR2* revealed no pathogenic sequence variation.[8,15,16] Data on the clinical phenotypes were collected from medical records or during physical examinations by one of the authors. Informed consent was obtained from patients and family members, and the study was approved by the responsible local ethics committee.

### Multiplex ligation-dependent probe amplification

MLPA was performed using 100 ng template DNA (referred to us or extracted from blood, tissue, or fibroblast samples) and the MLPA kit P155 (MRC-Holland, Amsterdam, The Netherlands), which contains MLPA probes for 10 of the 51 *COL3A1* exons, according to the manufacturer's instructions. MLPA fragments were separated by capillary electrophoresis on an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems, Rotkreuz, Switzerland). Each MLPA signal was normalized and compared with the corresponding peak area obtained in control DNA samples. Deviations of >30% were suspected as alterations and verified by repeated MLPA analysis.

### High-density microarray analyses

To narrow down the breakpoints of the hemizygous deletion identified by MLPA in patient 53B, high-density microarray analyses were performed using the GeneChip Human Mapping 500 K Array Set (Affymetrix, Santa Clara, CA, USA) according to the manufacturer's instructions. Data analysis was performed as described elsewhere.[16]

### Breakpoint analyses

Based on decreased microarray signal intensities, primers flanking the predicted deletion were designed (P53B_7F 5′-AAAAATAGGGCAATGTCAACTAA-3′, P53B_17R 5′-CTCGACCAGCTTCAGAACT-3′) and used in long-range PCR. Accordingly, the Expand Long Template PCR System (Roche Diagnostics, Rotkreuz, Switzerland) was used with 100 ng of DNA, Buffer 3, and thermal cycling program (annealing for 30 s at 58 °C and elongation for 15 min at 68 °C) as described previously.[16] Amplification products were sequenced using internal primers (P53B_7c_F 5′-GCAACAATGAATGGGAGAGA-3′, P53B_16c_R 5′-ACTCTGAATCAGCACCACTTG-3′) and standard procedures.[16] Family members were tested by both MLPA and a PCR-based assay using primers designed for sequencing of the deletion breakpoints.

### Biochemical testing and electron microscopy

Cultured dermal fibroblasts from the index patient 53B, his mother (53D), and two of his affected brothers (53 and 53E) were radiolabeled. Subsequently, collagens in medium and cell layer (harvested separately) were treated with pepsin, precipitated with ethanol, separated by SDS-PAGE (5%), and visualized by fluorography.[17] In addition, for patients 53, 53D, and 53E, a portion of the skin biopsies was processed for transmission electron microscopy as reported.[18] Furthermore, for patient 53E, his mother (53D), his daughter (53I), and his son (53J) as well as for his wife (53M), as a control subject, standard laboratory blood parameters thought to be associated with hemochromatosis type 4 were determined (for details, see Supplementary Table S2).

### Statistical analysis

For proportions, the upper and lower limits of the 95% confidence interval were calculated.[15]

## RESULTS

### Multiplex ligation-dependent probe amplification

We had screened 100 *FBN1*-, *TGFBR1*-, and *TGFBR2*-mutation-negative AD patients for large *COL3A1* deletions/duplications by MLPA. In one of these patients (53B), the relative peak areas of all MLPA probes for *COL3A1* were reduced, suggesting a hemizygous deletion of the entire gene (Supplementary Figure S2). The remaining 99 patients showed relative peak areas within the normal range. As the MLPA kit used in this study contains only probes for 10 of the 51 exons of *COL3A1* (exons 1, 2, 5, 11, 17, 20, 29, 32, 43, and 51; NM_000090.3), in the 99 negatively tested patients only loss or gain of these 10 exons could be excluded. In addition, mosaicisms and copy-number neutral rearrangements cannot be excluded as such cases may not be detectable by MLPA. Thus, in patient cohorts comparable to that used in this study the relative frequency of deletions and duplications of the 10 analyzed *COL3A1* exons can be expected as 1/100 (0.1–6.2%, $P$=0.05) and 0/100 (0.0–4.6%, $P$=0.05), respectively.

### Identification and characterization of breakpoints

Loss of heterozygosity and decreased signal intensities upon high-density microarray analyses confirmed the MLPA result of patient 53B and narrowed down the deletion breakpoints between rs16829183 and rs932169 (Figure 1b). Accordingly, primers were designed and used for long-range PCR. Because of the distance between the primer-binding sites on the normal allele (~3.41 Mb), only the amplification of the deletion-carrying allele was possible and resulted in an amplicon of ~8.5 kb (data not shown). Subsequent sequencing of this amplicon identified a deletion of 3 408 306 bp at 2q32.1q32.3 (Figure 1).

At the start and end points of the deletion, there is a short stretch of identical sequences (Figure 1c). This phenomenon has also been reported for other deletions.[16,19] One could speculate that the deletion presented here can be the result of mechanisms mainly responsible for copy number variations (CNVs) in the human genome, such as nonallelic homologous recombination (NAHR), non-homologous end joining (NHEJ), and Fork Stalling and Template Switching (FoSTeS).[20]

The deletion includes regions with known CNVs (Figure 1a and Supplementary Table S3). Although to date little is known about the conservation and effect of CNVs, the CNVs within the deleted region have been found in healthy individuals and thus considered to be non-pathogenic. This interpretation is also supported by the fact that all known CNVs in the deleted region are relatively small (<200 kb, cf. pathogenic CNVs are mostly >500 kb).[21] As expected, microarray analyses of patient 53B also revealed known CNVs in other parts of his genome (data not shown).

Furthermore, the deletion presented here affects not only *COL3A1* but also 21 other known genes (Figure 1a). Although the function of most of these genes is unknown/unclear, mutations in three genes (*COL5A2*, *SLC40A1*, and *MSTN*) have previously been associated with autosomal dominant disorders. Accordingly, mutations in *COL5A2* (MIM *120190), which encodes the α2 chain of type V collagen, lead to the classical type of EDS (EDS I/II, MIM #130000/#130010), the major diagnostic criteria of which include skin hyperextensibility, widened atrophic scars, and joint hypermobility but no AD.[14,22] The *SLC40A1* gene (*SLC11A3*, MIM *604653) encodes ferroportin and has been associated with hemochromatosis type 4 (HFE4, MIM #606069), a disorder of iron homeostasis.[23,24] *MSTN* (MIM +601788) encodes

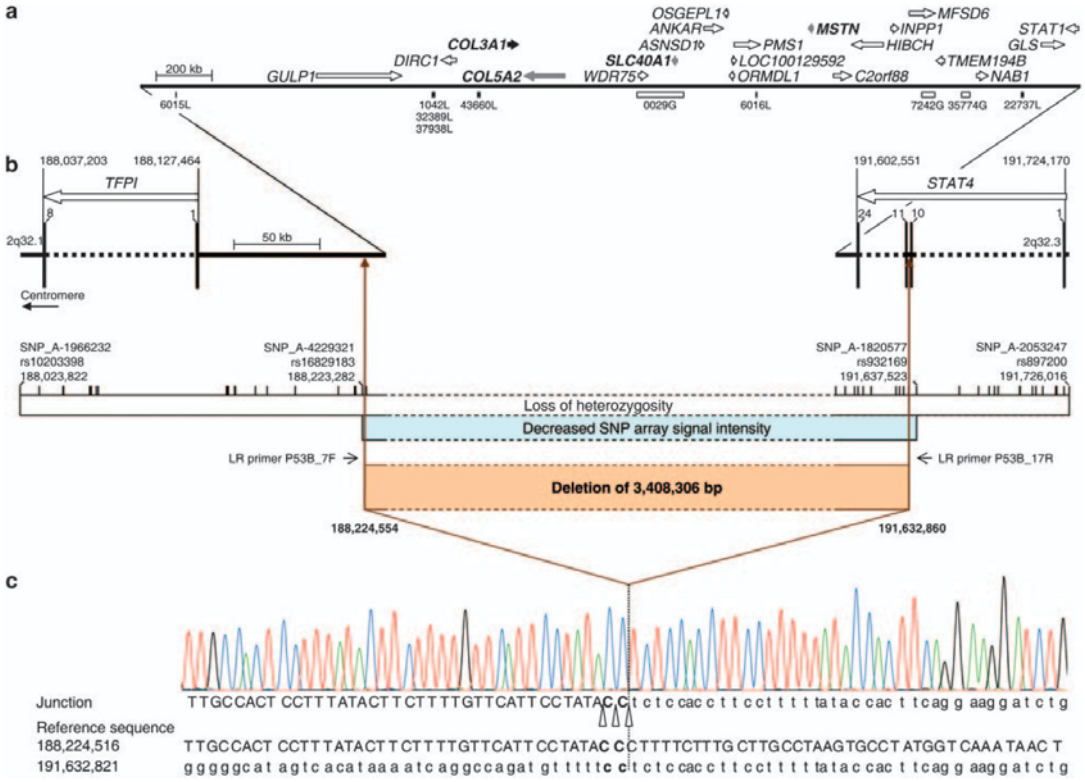



**Figure 1** Deletion of 3.4 Mb identified in this study. (**a**) Schematic representation of the completely deleted genes. Genes are represented by arrows that indicate the direction of transcription according to Entrez Gene (www.ncbi.nlm.nih.gov/sites/entrez?db=gene; version May 2009). *COL3A1* is denoted by a black arrow and other genes associated with a dominantly inherited disease and thus considered for further analyses are indicated by gray arrows. Known copy number variations (CNVs) in this region are given below the line according to the Database of Genomic Variants (DGV, http://projects.tcag.ca/variation, version May 2009; Supplementary Table S3). The black bars represent losses (L) and white bars gains (G). (**b**) Schematic representation of the region flanking the breakpoints. Exons are specified as bars and marked with the corresponding number. Regions derived from high-density microarray analyses are represented by a white bar for loss of heterozygosity and by a blue bar for decreased SNP array signal intensity. The positions of SNPs tested by the array set are indicated by vertical lines. The deleted region is denoted as a brown bar and the primers used for long-range (LR) PCR (LR primers P53B_7F and P53B_17R) are indicated by arrows. (**c**) Sequence of the long-range PCR product spanning the breakpoint junction of the deletion. Uppercase letters represent the sequence in the region of the start point of the deletion and lowercase letters the sequence in the region of the deletion end point. Because of identical sequences at the site of the breakpoints, the break and rejoining could have occurred at three positions as indicated by open triangles. The dotted line marks the most telomeric position of the possible breakpoints. All nucleotide positions are given in relation to the human genome reference sequence (NCBI build 36.1, March 2006).

the muscle growth inhibitor myostatin and has been found to be mutated in incomplete autosomal dominant muscle hypertrophy (MIM +601788).[25] We considered these three genes for further analyses in addition to *COL3A1*.

### Clinical and biochemical phenotypes

Most members of family 53 inherited the deleted allele (Figure 2a). Data on the clinical phenotype of the index patient 53B (Figure 2a) were collected from medical records. Accordingly, 53B died unexpectedly at the age of 34 years because of an abdominal aortic dissection found to be cranial from the celiac artery encompassing all segments of the descending thoracic aorta (Table 1). He was described as being in good general health but with prominent and early-onset varicose veins. In addition to patient 53B, his oldest brother (53, Figure 2a), who was also affected, was the only family member with reported aortic dissection, which occurred at the age of 43, 48, and 51 years, the latter leading to death (Table 1). Autopsy of patient 53 revealed not only the rupture of the thoracic aorta but also the involvement of medium-sized arteries.

In contrast, six familial carriers of the 3.4-Mb deletion (53, 53D, 53E, 53H, 53I, and 53J) and one non-carrier (53F), who served as an intrafamilial control subject, were available for physical examinations, which were focused on EDS IV, EDS I/II, muscle hypertrophy, and HFE4 (Figure 2, Table 1, and Supplementary Table S2). From the clinical features of EDS IV, all but one of the deletion carriers showed fragile and thin/translucent skin (Table 1). Varicose veins (Figure 2d) were detected only in adults, whereas hypermobility of small joints was predominantly present in younger deletion carriers. Facial acrogeria was limited to 53 and 53H, whereas acrogeria of hands and feet was present in all but 53I and 53J. Moderately thin lips were also

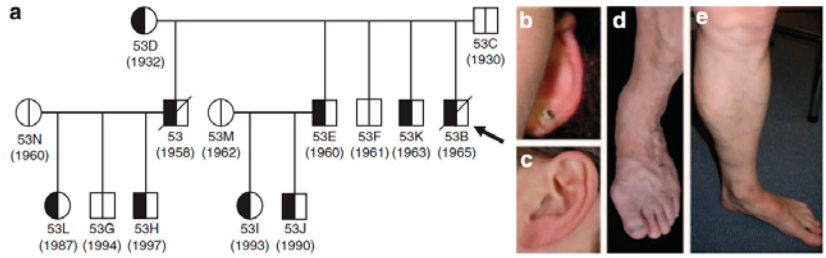



**Figure 2** Pedigree and some clinical features of Family 53. (**a**) Pedigree: the arrow denotes the male index patient (53B). The vertical line in the symbols (circle, female; square, male) denotes molecular genetic testing for the 3.4-Mb deletion: white halves represent normal alleles and black ones the allele carrying the deletion. The diagonal line through a symbol indicates deceased family members. The year of birth is given in parentheses. (**b**) Overfolded ear helix (53I). (**c**) Small earlobe (53). (**d**) Prominent varicose veins (53). (**e**) Muscle hypertrophy of the lower leg (53E).

**Table 1 Overview of clinical features in Family 53 regarding genes affected by the 3.4-Mb deletion identified in this study**

| *Clinical features* | *53* | *53D* | *53B* | *53E* | *53H* | *53I* | *53J* | *53F* |
|---|---|---|---|---|---|---|---|---|
| Deletion 2q32.1q32.3 (3.4 Mb) | + | + | + | + | + | + | + | – |
| Age at clinical examination (years) | 50 | 77 | 34 | 49 | 12 | 15 | 18 | 47 |
| Age at death due to arterial rupture (years) | 51 | N/A | 34 | N/A | N/A | N/A | N/A | N/A |
| *COL3A1 (EDS IV)* | | | | | | | | |
| Thin/translucent skin | + | + | + | + | + | + | – | – |
| Arterial/intestinal organ rupture | +[a,b,c] | – | +[d] | +[e] | – | – | – | – |
| Easy bruising | ++ | – | + | – | + | – | – | – |
| Early-onset varicose veins | ++ | ++ | ++ | ++ | – | – | – | – |
| Hypermobility of small joints | – | – | + | – | + | + | + | – |
| Missing ear lobe/ear helix abnormal | + | – | N/A | – | – | + | + | – |
| *COL5A2 (EDS I/II)* | | | | | | | | |
| Skin hyperextensibility | – | – | – | +[f] | – | – | – | – |
| Joint hypermobility | – | – | + | – | ++ | + | + | – |
| Smooth, velvety skin | + | – | + | – | + | – | – | – |
| Widened atrophic scars | – | – | – | – | – | – | – | – |
| Muscular hypotonia/delayed gross motor development | – | – | – | – | – | – | – | – |
| *MSTN (muscle hypertrophy)* | | | | | | | | |
| Increased muscle strength | + | – | N/A | + | + | – | + | – |
| Increased muscle size (gastrocnemius, soleus) | ++ | – | N/A | ++ | ++ | – | ++ | – |

Abbreviations: N/A, measurement or information not available or not applicable; –, not present; +, present; ++, strongly present.
[a]Dissection of infradiaphragmal infrarenal aorta.
[b]Dissection of arteria mesenterica superior.
[c]Dissection/rupture of thoracic aorta.
[d]Dissection/rupture of juxtarenal aorta.
[e]Rupture of bladder.
[f]Noticeable skin extensibility during skin biopsy (only).
See Figure 2a for pedigree and Supplementary Table S2 for laboratory blood parameters.

observed (53E, 53I, and 53J). Examination for major clinical signs of EDS I/II revealed smooth velvet skin and joint hypermobility, but none of the other classical clinical signs of EDS I/II were observed (Table 1). Impressively, all investigated male deletion carriers showed increase in muscle size of lower extremities with slightly increased muscle power, a finding consistent with myostatin-related muscle hypertrophy (Figure 2e and Table 1).

Apart from severe (early onset) varicosis and thin/translucent skin, the oldest deletion carrier (53D) is asymptomatic (Table 1). This could be explained by modifying gene(s)/factor(s) or by a mosaic of normal and mutated cells. The latter cannot be excluded, even if analysis of her leukocyte, saliva, and fibroblast DNAs provided no evidence for mosaicism (data not shown).

In addition, standard biochemical blood parameters determined for four deletion carriers (53D, 53E, 53J, and 53I) were comparable with normal values for age and sex (Supplementary Table S2), providing no evidence for HFE4. Notably, although HFE4 has previously been associated with heterozygous mutations in the *SLC40A1* gene,[23,24] in these four deletion carriers (53D, 53E, 53J, and 53I) without long-term low iron intake we found neither an increase in serum ferritin nor an abnormal transferrin saturation. In the 15-year-old female deletion carrier (53I), the relative number of hypochrome erythrocytes was slightly increased, but she did not have anemia and all other erythrocyte parameters were within normal range.

Taken together, despite (age related) clinical variability, deletion carriers showed pronounced clinical signs of EDS IV and muscle

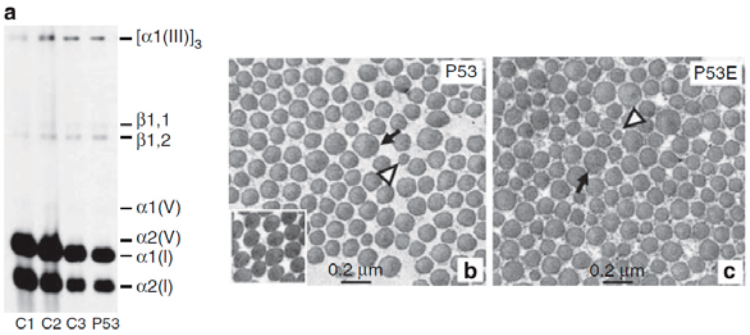



**Figure 3** Results of biochemical collagen analysis and electron microscopy. (**a**) SDS-PAGE of collagens produced by cultured dermal fibroblasts showing normal distribution and electrophoretic mobility of collagen type I (α1(I) and α2(I) chains), collagen type III (homotrimers of α1(III) chains), and collagen type V (α1(V) and α2(V) chains) in patient 53 (P53) and controls (C1, C2, and C3). Note that the α1(V) and α2(V) chains can only be seen on gels exposed for longer time. (**b, c**) Cross-sections of the dermis of patients 53 (P53) and 53E (P53E) showing abnormal collagen fibrils with either enlarged, irregular outlines (black arrows) or with abnormally small diameters (white arrowheads), occurring within otherwise regularly shaped fibrils. The insert in (**b**) shows a normal control.

hypertrophy, but only moderate expression of EDS I/II, resulting in a mixed clinical phenotype of these disorders (Table 1). No other clinical signs and symptoms were observed and no evidence of other disease (eg, cancer) by history was found in deletion carriers.

### Biochemical collagen analysis and electron microscopy

In deletion carriers 53, 53B, 53D, and 53E, we detected on SDS-PAGE a normal distribution as well as normal electrophoretic migration patterns for collagens I, III, and V in both medium and cell layer, confirming the normal function of the non-deleted *COL3A1* and *COL5A2* alleles (patient 53 in Figure 3a). In addition, our biochemical (protein based) collagen analyses confirmed the previous observation that such *in vitro* testing is less sensitive in identifying mutations that decrease production but do not alter the structure of type III and V collagens (Figure 3a).[26,27]

Electron micrographs of the dermis of patients 53, 53D, and 53E showed collagen fibrils with abnormally large diameters and slightly irregular outlines as well as abnormally small collagen fibril diameters, which were interspersed with normal-appearing fibrils (Figures 3b and c). The size variability of fibril diameters was increased compared with the control (cf. insert in Figure 3b). Longitudinal sections of fibrils consisted mostly of normally aggregated filaments as well as of few poorly aggregated ones lacking the typical transversal periodicity (data not shown). All these findings are consistent with a mixed phenotype of EDS IV and EDS I/II.

## DISCUSSION

We assessed the contribution of large *COL3A1* deletions/duplications to AD in patients in whom genetic testing of *FBN1*, *TGFBR1*, and *TGFBR2* revealed no mutation. We identified a hemizygous deletion of 3.4 Mb affecting *COL3A1* and other genes. Hence, we analyzed the clinical and biochemical effects of the hemizygous deletion of four of these genes (*COL3A1*, *COL5A2*, *MSTN*, and *SLC40A1*), each of which has previously been associated with an autosomal dominant disorder. Our data show that the hemizygous deletion of *COL3A1*, *COL5A2*, and *MSTN*, but not that of *SLC40A1*, leads to a clinical phenotype, extending the molecular etiology of EDS IV, EDS I/II, muscle hypertrophy, and HFE4, respectively.

In autosomal dominant disorders, haploinsufficiency can occur when a gene has only a single functional copy, instead of two copies. A distinction is drawn between true and functional haploinsufficiency. True haploinsufficiency is the *a priori* result of a hemizygous deletion, whereas functional haploinsufficiency occurs when one allele loses functionality, for example, because a key residue is mutated or the transcript amount of the mutant allele is reduced by NMD. As even in the era of CNVs, reports on true haploinsufficiency are rare, functional haploinsufficiency has often been used to assess whether or not haploinsufficiency of a gene is sufficient to cause features of the underlying disease.

*A priori*, the hemizygous deletion identified in this study causes true haploinsufficiency of the deleted genes (cf. no transcription of the deleted allele). The only question is which of these true haploinsufficiencies is responsible for the clinical phenotype. Thus, one would expect reduced (50%) expression of the affected genes. Indeed, our preliminary quantitative transcript analyses indicated reduction of *COL3A1* and *COL5A2* transcripts in cultured fibroblasts of deletion carriers (data not shown). However, cell culture conditions as well as modifying gene(s)/factor(s) can influence the expression of the undeleted allele, which can lead to expression levels significantly differing from 50%.[28] Furthermore, there is increasing evidence that CNVs can have intra- and inter-chromosomal effects on other genes because of interactions between chromosomal regions.[29] Thus, it is possible that the 3.4-Mb deletion identified in this study affects not only the expression of the deleted genes but also other genes on chromosome 2 and/or elsewhere in the genome. Similarly, modifying genetic, epigenetic, environmental, and stochastic factors can be the reason for the (age related) variability in the clinical phenotype of Family 53 (Table 1). For these reasons, we performed physical and biochemical examinations rather than *in vitro* transcript analyses of the deleted genes.

In previous studies, most of the identified *COL3A1* mutations causing EDS IV were missense or splice site mutations, leading to structural alterations of the protein. Only few cases of functional *COL3A1* haploinsufficiency, due to NMD, have been reported.[27,30] However, it has remained unclear whether or not the potent dominant-negative effect of the remaining mutant transcripts, which escape NMD, leads to the disease (cf. NMD is almost always incomplete). A recent study on homozygosity for a *COL3A1* null allele with NMD (p.Lys161GlnfsX45) could also not resolve this dilemma.[30] Cytogenetically detectable 2q32 deletions affect too many genes in addition

to *COL3A1*, leading to severe clinical phenotypes with developmental delay and thus hampering the clinical identification of EDS IV (Supplementary Figure S1 and Supplementary Table S1). In animal models, mice homozygous for an inactivated allele of *Col3a1* showed a phenotype closely resembling EDS IV, whereas heterozygous mice were phenotypically normal.[31] Late-onset signs in heterozygous mice, however, could not be excluded, because these mice had a limited follow-up period (cf. there are functional changes in bladder tissue of 8-week-old mice heterozygous for a *Col3a1* null allele).[32] Thus, this study is the first to show at the molecular level that the complete loss of one *COL3A1* allele (true haploinsufficiency) can cause an EDS-IV-related phenotype.

Similarly, no cases of true haploinsufficiency of *COL5A2* and *MSTN* have previously been described at the molecular level for EDS I/II and muscle hypertrophy, respectively, the clinical signs of which were clearly shown in this study. In comparison, both true and functional haploinsufficiencies of *COL5A1* (MIM *120215), the other gene associated with EDS I/II coding for the α1 chain in the $[\alpha1(V)]_2$, α2(V) heterotrimers of type V collagen, have been reported.[33–36] This difference in the haploinsufficiencies of *COL5A2* and *COL5A1* may be because of the notion/evidence that homotrimeric formation of α2(V) chains is not possible, whereas α1(V) chains can assemble into stable homotrimers (ie, *COL5A1* can replace *COL5A2*).[37,38] This may have hampered the identification of cases with *COL5A2* haploinsufficiency because of lack of sufficient clinical signs. Although AD may occur in EDS I/II,[39] none of the few (~10) *COL5A2* mutations reported so far have been associated with AD. In addition, mice with homozygous *Col5a2* deletion survived poorly, possibly because of complications from spinal (but not from cardiovascular) deformities, and showed skin and eye abnormalities as a result of disorganized type I collagen fibrils.[40] Consequently, the AD phenotype in Family 53 is most likely caused mainly by true haploinsufficiency of *COL3A1*; the haploinsufficiency of *COL5A2* may have merely a modifier effect.

In contrast, we found no evidence for HFE4 in four cases of true *SLC40A1* haploinsufficiency, not even in the 77-year-old carrier (cf. normal blood parameters; no signs for joint pains, osteoarthritis, fatigue, cardiomyopathies, and endocrine disorders). This lack of evidence for HFE4 is somewhat unexpected, as heterozygous loss-of-function *SLC40A1* mutations have previously been associated with HFE4.[41,42] However, as the age of onset of HFE4 is up to 60 years in males and ~10 years later in females (age-related penetrance),[43] it is possible that young carriers of the *SLC40A1* deletion failed to present clinical symptoms of HFE4 in this study because of their age of 15 (53I, female), 18 (53J, male), and 49 (53E, male) years at examination. As, to our knowledge, no patient with complete hemizygous deletion of this gene has previously been reported at the molecular level, this is the first description of the (so far normal) clinical phenotype of individuals with true *SLC40A1* haploinsufficiency, suggesting the negligible role of haploinsufficiency in the pathogenesis of HFE4.

Altogether, the hemizygous deletion presented here causes a mixed phenotype because of the true haploinsufficiencies of *COL3A1*, *COL5A2*, and *MSTN*, whereby most likely the true haploinsufficiency of *COL3A1* causes AD. In the case of *SLC40A1*, we provided evidence that a haploinsufficient gene known to be associated with an autosomal dominant disorder does not necessarily lead to a clinical phenotype. Deleted genes that were not further investigated in this study have been associated with a recessive disease or the gene function, and/or their disease association is unknown/unclear. As no evidence of other clinical signs, symptoms, and diseases (eg, cancer) was found in this study, the effect of the haploinsufficiency of these other genes located within the deletion (eg, *PMS1*) is low, if any. Our finding that true haploinsufficiency leads to an EDS-IV-related phenotype opens the possibility of novel therapeutic strategies, which increase expression of the undeleted allele(s). In addition, this work emphasizes the inclusion of deletion/duplication screening in the comprehensive genetic testing of *COL3A1* as well as the importance of this testing in AD patients, at least with EDS-like phenotypes.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGEMENTS

We are grateful to Family 53, all patients, and referring physicians for participating in this study. We thank Matthias Baumgartner, Wolfgang Berger, and Albert Schinzel for their support; Philippe Reuge for help with MLPA analyses; Angelika Schwarze for help with cell cultures and biochemical collagen analyses; and Michal Okoniewski, Elisabeth Probst, and members of the Institute of Medical Genetics, University of Zurich, for discussions. This work was supported by the FGCZ and grants from the Swiss National Science Foundation (3100A0–120504 to GM and 3200B0–109370/1 to BS).

1 Pyeritz RE: The Marfan syndrome. *Annu Rev Med* 2000; **51**: 481–510.
2 De Paepe A, Devereux RB, Dietz HC, Hennekam RCM, Pyeritz RE: Revised diagnostic criteria for the Marfan syndrome. *Am J Med Genet* 1996; **62**: 417–426.
3 Dietz HC, Cutting GR, Pyeritz RE *et al*: Marfan syndrome caused by a recurrent *de novo* missense mutation in the fibrillin gene. *Nature* 1991; **352**: 337–339.
4 Mizuguchi T, Collod-Beroud G, Akiyama T *et al*: Heterozygous TGFBR2 mutations in Marfan syndrome. *Nat Genet* 2004; **36**: 855–860.
5 Loeys BL, Chen J, Neptune ER *et al*: A syndrome of altered cardiovascular, craniofacial, neurocognitive and skeletal development caused by mutations in TGFBR1 or TGFBR2. *Nat Genet* 2005; **37**: 275–281.
6 Pannu H, Fadulu VT, Chang J *et al*: Mutations in transforming growth factor-beta receptor type II cause familial thoracic aortic aneurysms and dissections. *Circulation* 2005; **112**: 513–520.
7 Loeys BL, Schwarze U, Holm T *et al*: Aneurysm syndromes caused by mutations in the TGF-beta receptor. *N Engl J Med* 2006; **355**: 788–798.
8 Matyas G, Arnold E, Carrel T *et al*: Identification and in silico analyses of novel TGFBR1 and TGFBR2 mutations in Marfan syndrome-related disorders. *Hum Mutat* 2006; **27**: 760–769.
9 Germain DP: Ehlers-Danlos syndrome type IV. *Orphanet J Rare Dis* 2007; **2**: 32.
10 Beighton P, De Paepe A, Steinmann B, Tsipouras P, Wenstrup RJ: Ehlers-Danlos syndromes: revised nosology, Villefranche, 1997. Ehlers-Danlos National Foundation (USA) and Ehlers-Danlos Support Group (UK). *Am J Med Genet* 1998; **77**: 31–37.
11 Pope FM, Martin GR, Lichtenstein JR *et al*: Patients with Ehlers-Danlos syndrome type IV lack type III collagen. *Proc Natl Acad Sci USA* 1975; **72**: 1314–1316.
12 Superti-Furga A, Gugler E, Gitzelmann R, Steinmann B: Ehlers-Danlos syndrome type IV: a multi-exon deletion in one of the two COL3A1 alleles affecting structure, stability, and processing of type III procollagen. *J Biol Chem* 1988; **263**: 6226–6232.
13 Tromp G, Kuivaniemi H, Shikata H, Prockop DJ: A single base mutation that substitutes serine for glycine 790 of the alpha 1 (III) chain of type III procollagen exposes an arginine and causes Ehlers-Danlos syndrome IV. *J Biol Chem* 1989; **264**: 1349–1352.
14 Steinmann B, Royce PM, Superti-Furga A: The Ehlers-Danlos syndrome. in: Royce PM, Steinmann B (eds): *Connective Tissue and its Heritable Disorders*. New York: Wiley-Liss, 2002; 431–523.
15 Matyas G, De Paepe A, Halliday D, Boileau C, Pals G, Steinmann B: Evaluation and application of denaturing HPLC for mutation detection in Marfan syndrome: identification of 20 novel mutations and two novel polymorphisms in the FBN1 gene. *Hum Mutat* 2002; **19**: 443–456.
16 Matyas G, Alonso S, Patrignani A *et al*: Large genomic fibrillin-1 (FBN1) gene deletions provide evidence for true haploinsufficiency in Marfan syndrome. *Hum Genet* 2007; **122**: 23–32.
17 Steinmann B, Rao VH, Vogel A, Bruckner P, Gitzelmann R, Byers PH: Cysteine in the triple-helical domain of one allelic product of the alpha 1(I) gene of type I collagen produces a lethal form of osteogenesis imperfecta. *J Biol Chem* 1984; **259**: 11129–11138.
18 Vogel A, Holbrook KA, Steinmann B, Gitzelmann R, Byers PH: Abnormal collagen fibril structure in the gravis form (type I) of Ehlers-Danlos syndrome. *Lab Invest* 1979; **40**: 201–206.
19 Giacalone JP, Francke U: Common sequence motifs at the rearrangement sites of a constitutional X/autosome translocation and associated deletion. *Am J Hum Genet* 1992; **50**: 725–741.
20 Gu W, Zhang F, Lupski JR: Mechanisms for human genomic rearrangements. *Pathogenetics* 2008; **1**: 4.
21 Bruno DL, Ganesamoorthy D, Schoumans J *et al*: Detection of cryptic pathogenic copy number variations and constitutional loss of heterozygosity using high resolution SNP

microarray analysis in 117 patients referred for cytogenetic analysis and impact on clinical practice. *J Med Genet* 2009; **46**: 123–131.
22 Michalickova K, Susic M, Willing MC, Wenstrup RJ, Cole WG: Mutations of the alpha2(V) chain of type V collagen impair matrix assembly and produce Ehlers-Danlos syndrome type I. *Hum Mol Genet* 1998; **7**: 249–255.
23 Montosi G, Donovan A, Totaro A *et al*: Autosomal-dominant hemochromatosis is associated with a mutation in the ferroportin (SLC11A3) gene. *J Clin Invest* 2001; **108**: 619–623.
24 Njajou OT, Vaessen N, Joosse M *et al*: A mutation in SLC11A3 is associated with autosomal dominant hemochromatosis. *Nat Genet* 2001; **28**: 213–214.
25 Schuelke M, Wagner KR, Stolz LE *et al*: Myostatin mutation associated with gross muscle hypertrophy in a child. *N Engl J Med* 2004; **350**: 2682–2688.
26 Pepin M, Schwarze U, Superti-Furga A, Byers PH: Clinical and genetic features of Ehlers-Danlos syndrome type IV, the vascular type. *N Engl J Med* 2000; **342**: 673–680.
27 Schwarze U, Schievink WI, Petty E *et al*: Haploinsufficiency for one COL3A1 allele of type III procollagen results in a phenotype similar to the vascular form of Ehlers-Danlos syndrome, Ehlers-Danlos syndrome type IV. *Am J Hum Genet* 2001; **69**: 989–1001.
28 Hutchinson S, Furger A, Halliday D *et al*: Allelic variation in normal human FBN1 expression in a family with Marfan syndrome: a potential modifier of phenotype? *Hum Mol Genet* 2003; **12**: 2269–2276.
29 Henrichsen CN, Vinckenbosch N, Zollner S *et al*: Segmental copy number variation shapes tissue transcriptomes. *Nat Genet* 2009; **41**: 424–429.
30 Plancke A, Holder-Espinasse M, Rigau V, Manouvrier S, Claustres M, Van Kien PK: Homozygosity for a null allele of COL3A1 results in recessive Ehlers-Danlos syndrome. *Eur J Hum Genet* 2009; **17**: 1411–1416.
31 Liu X, Wu H, Byrne M, Krane S, Jaenisch R: Type III collagen is crucial for collagen I fibrillogenesis and for normal cardiovascular development. *Proc Natl Acad Sci USA* 1997; **94**: 1852–1856.
32 Stevenson K, Kucich U, Whitbeck C, Levin RM, Howard PS: Functional changes in bladder tissue from type III collagen-deficient mice. *Mol Cell Biochem* 2006; **283**: 107–114.
33 Toriello HV, Glover TW, Takahara K *et al*: A translocation interrupts the COL5A1 gene in a patient with Ehlers-Danlos syndrome and hypomelanosis of Ito. *Nat Genet* 1996; **13**: 361–365.
34 Wenstrup RJ, Florer JB, Willing MC *et al*: COL5A1 haploinsufficiency is a common molecular mechanism underlying the classical form of EDS. *Am J Hum Genet* 2000; **66**: 1766–1776.
35 Bouma P, Cabral WA, Cole WG, Marini JC: COL5A1 exon 14 splice acceptor mutation causes a functional null allele, haploinsufficiency of alpha 1(V) and abnormal heterotypic interstitial fibrils in Ehlers-Danlos syndrome II. *J Biol Chem* 2001; **276**: 13356–13364.
36 Mitchell AL, Schwarze U, Jennings JF, Byers PH: Molecular mechanisms of classical Ehlers-Danlos syndrome (EDS). *Hum Mutat* 2009; **30**: 995–1002.
37 Fichard A, Tillet E, Delacoux F, Garrone R, Ruggiero F: Human recombinant alpha1(V) collagen chain. Homotrimeric assembly and subsequent processing. *J Biol Chem* 1997; **272**: 30083–30087.
38 Chanut-Delalande H, Bonod-Bidaud C, Cogne S *et al*: Development of a functional skin matrix requires deposition of collagen V heterotrimers. *Mol Cell Biol* 2004; **24**: 6049–6057.
39 Wenstrup RJ, Meyer RA, Lyle JS *et al*: Prevalence of aortic root dilation in the Ehlers-Danlos syndrome. *Genet Med* 2002; **4**: 112–117.
40 Andrikopoulos K, Liu X, Keene DR, Jaenisch R, Ramirez F: Targeted mutation in the col5a2 gene reveals a regulatory role for type V collagen during matrix assembly. *Nat Genet* 1995; **9**: 31–36.
41 Wallace DF, Pedersen P, Dixon JL *et al*: Novel mutation in ferroportin1 is associated with autosomal dominant hemochromatosis. *Blood* 2002; **100**: 692–694.
42 Schimanski LM, Drakesmith H, Merryweather-Clarke AT *et al*: *In vitro* functional analysis of human ferroportin (FPN) and hemochromatosis-associated FPN mutations. *Blood* 2005; **105**: 4096–4102.
43 Cremonesi L, Forni GL, Soriani N *et al*: Genetic and clinical heterogeneity of ferroportin disease. *Br J Haematol* 2005; **131**: 663–670.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)

**Supplementary Table S1** Comparative overview of the clinical features of present cases and previously described patients carrying a cytogenetically detectable 2q32.2 deletion with a size of ~2q32 (~14.4 Mb) or less (blue marked in Supplementary Figure S1)

| | Deletion | Number of patients | Age (years) | Connective tissue rupture | Skin | Joints | Ears | Muscles | Developmental delay |
|---|---|---|---|---|---|---|---|---|---|
| Present cases (53, 53B, 53D, 53E, 53H, 53I, and 53J; see also Table 1, Figures 2 and 3) | 2q32.1q32.3 (3.4 Mb) | 7 | 12, 15, 18, 34, 49, 50, and 77 | arterial rupture (2/7)<br>rupture of bladder (1/7) | thin translucent skin (6/7)<br>smooth velvet skin (3/7)<br>easy bruising (3/7)<br>early onset varicose veins (4/7) | hypermobility of (small) joints (4/7) | missing ear lobe / abnormal ear helix (3/6) | increased muscle size and strength (4/6) | not present |
| de Ravel et al.[1] | 2q32.2q33.2 (~14.2 Mb) | 1 | 20 | no inguinal hernia | no subcutaneous fat<br>varicose veins of the lover limbs | not specified | low set | intially globally hypotonic, later peripherally hypertonic | mental retardation<br>delayed development |
| Prontera et al.[2] | 2q31.2q32.2 (~13.7 Mb) | 1 | 36 | inguinal hernia | deep palmar creases | normal | dysmorphic and retrorotated | peculiar "muscular build" (mainly legs and calves) | mental retardation<br>delayed development |
| Mencarelli et al.[3] | 2q31.2q32.3 (~12.7 Mb) | 1 | 14 | inguinal hernia | thin and transparent skin | normal | dysmorphic right ear (antihelix and extrafolds) | rigidity in the upper limbs | mental retardation<br>delayed development |
| Gallagher et al.[4] | 2q32.1q32.2 / 2q32.2q32.3 (~8.9 / 8 Mb) | 1 | 14 | not specified | not specified | hyperextensibility around the wrist | not specified | not specified | autism<br>delayed development |
| Kreuz and Wittwer[5] | 2q32 (~14.4 Mb) | 3 (mother and her two sons) | 2, 3, and 29 | umbilical hernia (2/3) | prominent veins<br>wrinkled skin<br>increased palmar/plantar creases<br>dry skin | laxity of joints (1/3) | not specified | muscle hypotonia (2/3) | mental retardation<br>delayed development |
| Glass et al.[6] | 2q32.2q33.1 (~14.4 Mb) | 1 | 16 | not specified | skin pigmentation | not specified | large or low set | not specified | mental retardation<br>delayed development |
| Pai et al.[7] | 2q32 (~14.4 Mb) | 2 (sisters) | infant (9 wk) and 3.5 | not specified | palmar creases faintly marked | hyperextensible joints | large slightly anteverted | normal | mental retardation<br>delayed development |


1 de Ravel TJ, Balikova I, Thiry P, Vermeesch JR, Frijns JP: Another patient with a de novo deletion further delineates the 2q33.1 microdeletion syndrome. *Eur J Med Genet* 2009; **52**: 120-122.
2 Prontera P, Bernardini L, Stangoni G *et al*: 2q31.2q32.3 deletion syndrome: report of an adult patient. *Am J Med Genet A* 2009; **149A**: 706-712.
3 Mencarelli MA, Caselli R, Pescucci C *et al*: Clinical and molecular characterization of a patient with a 2q31.2-32.3 deletion identified by array-CGH. *Am J Med Genet A* 2007; **143A**: 858-865.
4 Gallagher L, Becker K, Kearney G *et al*: Brief report: A case of autism associated with del(2)(q32.1q32.2) or (q32.2q32.3). *J Autism Dev Disord* 2003; **33**: 105-108.
5 Kreuz FR, Wittwer BH: Del(2q)--cause of the wrinkly skin syndrome? *Clin Genet* 1993; **43**: 132-138.
6 Glass IA, Swindlehurst CA, Aitken DA, McCrea W, Boyd E: Interstitial deletion of the long arm of chromosome 2 with normal levels of isocitrate dehydrogenase. *J Med Genet* 1989; **26**: 127-130.
7 Pai GS, Rogers JF, Sommer A: Identical multiple congenital anomalies/mental retardation (MCA/MR) syndrome due to del(2)(q32) in two sisters with intrachromosomal insertional translocation in their father. *Am J Med Genet* 1983; **14**: 189-195.

**Supplementary Table S2** Laboratory blood parameters in deletion-carrier grandmother (53D), father (53E), son (53J), and daughter (53I) as well as in deletion-noncarrier mother (53M, as control subject)

| Parameter | 53M (1962) | | | 53D (1932) | | | 53E (1960) | | | 53J (1990) | | | 53I (1993) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Result | Unit | Reference range | Result | Unit | Reference range | Result | Unit | Reference range | Result | Unit | Reference range | Result | Unit | Reference range |
| **Clinical chemistry** | | | | | | | | | | | | | | | |
| Total bilirubin | 6 | µmol/L | <17 | 7 | µmol/L | <21 | 9 | µmol/L | <17 | 11 | µmol/L | <17 | 10 | µmol/L | <17 |
| Creatinine | 63 | µmol/L | 44-80 | 76 | µmol/L | 44-80 | 76 | µmol/L | 62-106 | 80 | µmol/L | 62-106 | 53 | µmol/L | 44-80 |
| Creatine kinase | 107 | U/L | <167 | 163 | U/L | <167 | 98 | U/L | <190 | 134 | U/L | <190 | 78 | U/L | <167 |
| C-reactive protein | <1 | mg/L | <5 | <1 | mg/L | <5 | <1 | mg/L | <5 | <1 | mg/L | <5 | <1 | mg/L | <5 |
| Serum iron | 9.3 | µmol/L | 7.0-26.0 | 12.7 | µmol/L | 7.0-26.0 | 15 | µmol/L | 11.0-28.0 | 23.4 | µmol/L | 11.0-28.0 | 28.4 | µmol/L | 7.0-26.0 |
| Ferritin | 10 | µg/L | 10-150 | 85 | µg/L | 30-400 | 108 | µg/L | 30-400 | 40 | µg/L | 30-400 | 49 | µg/L | 20-200 |
| Transferrin | 38 | µmol/L | 25-50 | 35 | µmol/L | 25-50 | 27 | µmol/L | 25-50 | 31 | µmol/L | 25-50 | 36 | µmol/L | 25-50 |
| Transferrin saturation | 0.12 | | 0.15-0.50 | 0.18 | | 0.15-0.50 | 0.28 | | 0.20-0.55 | 0.38 | | 0.20-0.55 | 0.39 | | 0.15-0.50 |
| Soluble transferrin receptor | 4.22 | mg/L | 1.9-4.4 | 1.17 | mg/L | 0.76-1.76 | 2.68 | mg/L | 2.2-5.0 | 2.83 | mg/L | 2.2-5.0 | 2.88 | mg/L | 1.9-4.4 |
| **Hematology** | | | | | | | | | | | | | | | |
| Hemoglobin | 12.2 | g/dl | 11.7-15.3 | not performed | | | 14.7 | g/dl | 13.4-17.0 | 15.2 | g/dl | 13.4-17.0 | 13.5 | g/dl | 11.7-15.3 |
| Hematocrit | 36.0 | % | 35-46 | not performed | | | 43.3 | % | 40-50 | 43.4 | % | 40-50 | 40.2 | % | 35-46 |
| Red blood cell count | 4.09 | $10^6$/µl | 3.9-5.2 | not performed | | | 4.95 | $10^6$/µl | 4.2-5.7 | 5.22 | $10^6$/µl | 4.2-5.7 | 4.66 | $10^6$/µl | 3.9-5.2 |
| MCV | 88.1 | fl | 80-100 | not performed | | | 87.3 | fl | 80-100 | 83.0 | fl | 80-100 | 86.2 | fl | 80-100 |
| MCH | 29.9 | pg | 26-34 | not performed | | | 29.6 | pg | 26-34 | 29.0 | pg | 26-34 | 29.0 | pg | 26-34 |
| MCHC | 34.0 | g/dl | 31-36 | not performed | | | 33.9 | g/dl | 31-36 | 35.0 | g/dl | 31-36 | 33.7 | g/dl | 31-36 |
| Microcytes | 0.7 | % | 0-2.0 | not performed | | | 0.5 | % | 0-2.0 | 1.1 | % | 0-2.0 | 1.0 | % | 0-2.0 |
| Macrocytes | 0.3 | % | 0-2.0 | not performed | | | 0.2 | % | 0-2.0 | 0.1 | % | 0-2.0 | 0.1 | % | 0-2.0 |
| Hypochromic RBC | 1.4 | % | 0-2.0 | not performed | | | 0.8 | % | 0-2.0 | 0.5 | % | 0-2.0 | 2.8 | % | 0-2.0 |
| Hyperchromic RBC | 0.5 | % | 0-2.0 | not performed | | | 0.8 | % | 0-2.0 | 1.6 | % | 0-2.0 | 0.4 | % | 0-2.0 |
| Platelets | 297 | $10^3$/µl | 143-400 | not performed | | | 207 | $10^3$/µl | 143-400 | 182 | $10^3$/µl | 143-400 | 265 | $10^3$/µl | 143-400 |
| Left shift | 0 | | no (0) | not performed | | | 0 | | no (0) | 0 | | no (0) | 0 | | no (0) |
| White blood cell count | 4.48 | $10^3$/µl | 3.0-9.6 | not performed | | | 7.11 | $10^3$/µl | 3.0-9.6 | 5.88 | $10^3$/µl | 3.0-9.6 | 5.99 | $10^3$/µl | 3.0-9.6 |
| Neutrophils | 2.46 | $10^3$/µl | 1.40-8.00 | not performed | | | 4.51 | $10^3$/µl | 1.40-8.00 | 3.30 | $10^3$/µl | 1.40-8.00 | 3.25 | $10^3$/µl | 1.40-8.00 |
| Monocytes | 0.31 | $10^3$/µl | 0.16-0.95 | not performed | | | 0.55 | $10^3$/µl | 0.16-0.95 | 0.41 | $10^3$/µl | 0.16-0.95 | 0.40 | $10^3$/µl | 0.16-0.95 |
| Eosinophils | 0.11 | $10^3$/µl | 0.00-0.70 | not performed | | | 0.25 | $10^3$/µl | 0.00-0.70 | 0.46 | $10^3$/µl | 0.00-0.70 | 0.37 | $10^3$/µl | 0.00-0.70 |
| Basophils | 0.05 | $10^3$/µl | 0.00-0.15 | not performed | | | 0.03 | $10^3$/µl | 0.00-0.15 | 0.05 | $10^3$/µl | 0.00-0.15 | 0.05 | $10^3$/µl | 0.00-0.15 |
| Lymphocytes | 1.44 | $10^3$/µl | 1.50-4.00 | not performed | | | 1.60 | $10^3$/µl | 1.50-4.00 | 1.54 | $10^3$/µl | 1.50-4.00 | 1.77 | $10^3$/µl | 1.50-4.00 |
| LUC (non-classifiable lymphoblast cells) | 2.6 | % | 0.0-4.0 | not performed | | | 2.3 | % | 0.0-4.0 | 2.0 | % | 0.0-4.0 | 2.6 | % | 0.0-4.0 |
| **Immunological assays** | | | | | | | | | | | | | | | |
| Complement C3c | 0.74 | g/L | 0.9-1.8 | not performed | | | 0.83 | g/L | 0.9-1.8 | 0.85 | g/L | 0.9-1.8 | 0.89 | g/L | 0.9-1.8 |
| Complement C4 | 0.1 | g/L | 0.1-0.4 | not performed | | | 0.18 | g/L | 0.1-0.4 | 0.09 | g/L | 0.1-0.4 | 0.14 | g/L | 0.1-0.4 |
| Rheumatoid factor | <8 | IU/ml | <20 | not performed | | | <8 | IU/ml | <20 | <8 | IU/ml | <20 | <8 | IU/ml | <20 |
| ANA | negative | | <1:100 | not performed | | | negative | | <1:100 | negative | | <1:100 | negative | | <1:100 |
| Anti-CCP | 1 | U/ml | <30 | not performed | | | 0 | U/ml | <30 | 0 | U/ml | <30 | 0 | U/ml | <30 |

**Dietary habits:** 53D and 53E are no vegetarians, whereas 53I and 53J are vegetarians for the last ~1.5 years and ~6 months, respectively, but both eat fish. Due to the lack of long-term low iron intake, dietary habits of these deletion-carriers may not influence iron homeostasis. However, the effect of dietary habits on HFE4 has not been investigated.

**Clinical chemistry:** Venous plasma and serum samples were obtained by venous puncture and subsequent centrifugation of the blood samples. The concentrations of creatinine, total bilirubin, C-reactive protein (CRP), serum iron, and creatine kinase were assayed on a Roche/Hitachi Modular System P (Roche Diagnostics, Rotkreuz, Switzerland) according to the manufacturer's specifications and using proprietary reagents. The ferritin immunoassay was performed on the Modular Analytics E170 analyzer (Roche Diagnostics, Rotkreuz, Switzerland). Transferrin was measured on a COBAS Integra analyzer (Roche Diagnostics, Rotkreuz, Switzerland). Soluble transferrin receptor (sTfR) was analyzed using a nephelometric technique (Dade Behring, Marburg, Germany). Value of slightly elevated serum iron in 53I was not further investigated.

**Hematology:** Blood counts were analyzed using an Advia 2120 automated hematology system (Siemens Healthcare Diagnostics, Eschborn, Germany).

**Immunological assays:** Rheumatoid factor (RF) and complement factors C4 and C3c were determined on a Dade Behring nephelometer BNII (Dade Behring, Düdingen, Switzerland) according to the manufacturer's instructions (low C3c values are most likely due to immediate analysis; i.e. there was very short pre-analytical delay). Antinuclear antibodies (ANA) were determined by indirect immunofluorescence microscopy using Hep2-cells and rat tissues (Euroimmun Basis Profil 3B FB1805-2005-3). ANA titers less than 1:100 were considered negative. Antibodies to cyclic citrullinated peptide (Anti-CCP) were measured by Immunoscan RA anti-CCP ELISA (Euro-Diagnostica, Malmö, Sweden).

Immunological assays were performed to assess the impact of the partial deletion of *STAT4* (Figure 1), which has been associated with rheumatoid arthritis (RA, MIM #180300) and systemic lupus erythematosus (SLE, MIM #152700) (Supplementary Table S1). RF and Anti-CCP were analyzed regarding RA and ANA regarding SLE (anti-dsDNA and anti-Sm antibodies were not analyzed because ANA was negative). The analyzed parameters were in all three deletion-carriers within the normal range, providing no evidence for RA and SLE. This lack of association of the partial deletion of *STAT4* with RA and SLE is not inconsistent with current knowledge. However, RA and SLE cannot be completely excluded by these immunological assays.

**Supplementary Table S3** Known copy number variations (CNVs) larger than 1 kb within the deletion of 3.4 Mb identified in this study

| Name[a] | Type | Start position | End position | Size (bp) | Detection rate[b] | Known genes | Reference |
|---|---|---|---|---|---|---|---|
| 6015 | Loss | 188,360,896 | 188,363,219 | 2,324 | 1/36 | --- | Mills et al.[1] |
| 1042 | Loss | 189,277,306 | 189,287,290 | 9,985 | 1/30 | --- | Conrad et al.[2] |
| 32389 | Loss | 189,278,292 | 189,286,640 | 8,349 | 2/30 | --- | Perry et al.[3] |
| 37938 | Loss | 189,276,201 | 189,287,574 | 11,374 | 17/270 | --- | McCarroll et al.[4] |
| 43660 | Loss | 189,438,791 | 189,448,384 | 9,593 | 1/1 | --- | Wang et al.[5] |
| 0029 | Gain | 190,008,980 | 190,176,870 | 167,891 | 2/39 | *WDR75*, *SLC40A1* | Iafrate et al.[6] |
| 6016 | Loss | 190,434,734 | 190,437,628 | 2,895 | 1/36 | *PMS1* | Mills et al.[1] |
| 7242 | Gain | 191,027,036 | 191,073,955 | 46,920 | 2/50 | *MFSD6* | De Smith et al.[7] |
| 35774 | Gain | 191,169,323 | 191,199,654 | 30,332 | 1/1 | --- | Kidd et al.[8] |
| 22737 | Loss | 191,382,600 | 191,389,807 | 7,208 | 1/2 | --- | Korbel et al.[9] |

[a] According to the Database of Genomic Variants (DGV, http://projects.tcag.ca/variation, version May, 2009)
[b] Number of individuals harbouring the corresponding CNV in relation to the total number of individuals analyzed


1 Mills RE, Luttig CT, Larkins CE *et al*: An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Res* 2006; **16**: 1182-1190.

2 Conrad DF, Andrews TD, Carter NP, Hurles ME, Pritchard JK: A high-resolution survey of deletion polymorphism in the human genome. *Nat Genet* 2006; **38**: 75-81.

3 Perry GH, Ben-Dor A, Tsalenko A *et al*: The fine-scale and complex architecture of human copy-number variation. *Am J Hum Genet* 2008; **82**: 685-695.

4 McCarroll SA, Kuruvilla FG, Korn JM *et al*: Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat Genet* 2008; **40**: 1166-1174.

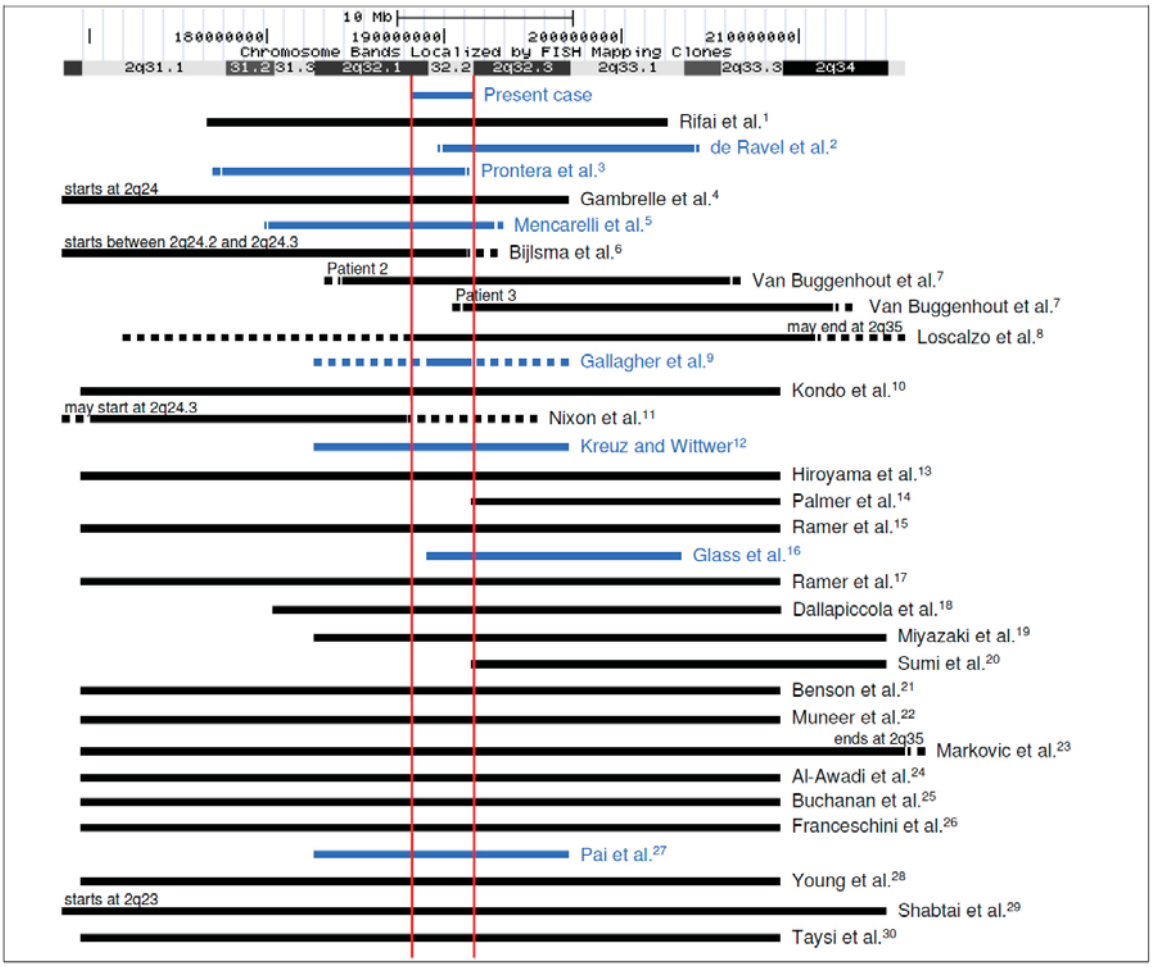
5 Wang J, Wang W, Li R *et al*: The diploid genome sequence of an Asian individual. *Nature* 2008; **456**: 60-65.

6 Iafrate AJ, Feuk L, Rivera MN *et al*: Detection of large-scale variation in the human genome. *Nat Genet* 2004; **36**: 949-951.

7 De Smith AJ, Tsalenko A, Sampas N *et al*: Array CGH analysis of copy number variation identifies 1284 new genes variant in healthy white males: implications for association studies of complex diseases. *Hum Mol Genet* 2007; **16**: 2783-2794.

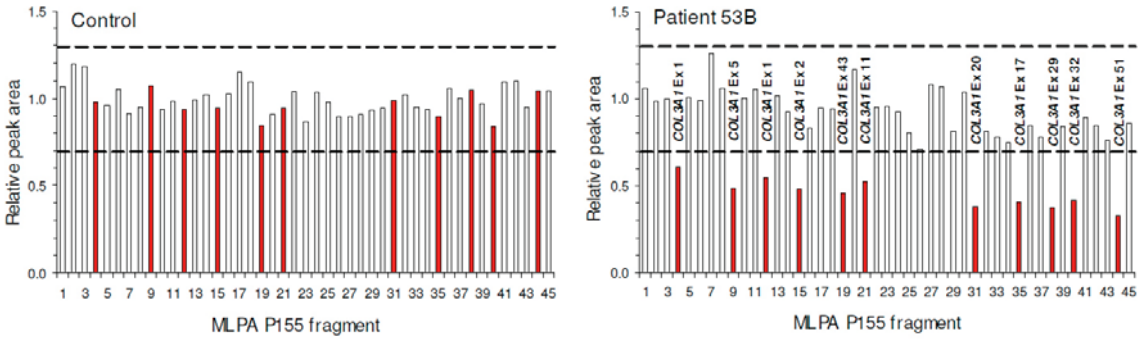
8 Kidd JM, Cooper GM, Donahue WF *et al*: Mapping and sequencing of structural variation from eight human genomes. *Nature* 2008; **453**: 56-64.

9 Korbel JO, Urban AE, Affourtit JP *et al*: Paired-end mapping reveals extensive structural variation in the human genome. *Science* 2007; **318**: 420-426.




**Supplementary Figure S1** Overview of previously reported cases with cytogenetically detectable deletions affecting the region deleted in this study (Present case). Deletions larger than del(2)(q23q35) are not included because such aberrations affect much larger chromosomal regions and thus much more genes. Cases (deletions <15 Mb) compared in Supplementary Table S1 are indicated by blue colour. The maximal possible extent of the deletions is indicated (dashed lines).


1 Rifai L, Port-Lis M, Tabet AC *et al*: Ectodermal dysplasia-like syndrome with mental retardation due to contiguous gene deletion: further clinical and molecular delineation of del(2q32) syndrome. *Am J Med Genet A* 2010; **152A**: 111-117.
2 de Ravel TJ, Balikova I, Thiry P, Vermeesch JR, Frijns JP: Another patient with a de novo deletion further delineates the 2q33.1 microdeletion syndrome. *Eur J Med Genet* 2009; **52**: 120-122.
3 Prontera P, Bernardini L, Stangoni G *et al*: 2q31.2q32.3 deletion syndrome: report of an adult patient. *Am J Med Genet A* 2009; **149A**: 706-712.
4 Gambrelle J, Till M, Lukusa B *et al*: Ocular anomalies associated with interstitial deletion of chromosome 2q31: case report and review. *Ophthalmic Genet* 2007; **28**: 105-109.
5 Mencarelli MA, Caselli R, Pescucci C *et al*: Clinical and molecular characterization of a patient with a 2q31.2-32.3 deletion identified by array-CGH. *Am J Med Genet A* 2007; **143A**: 858-865.
6 Bijlsma EK, Knegt AC, Bilardo CM, Goodman FR: Increased nuchal translucency and split-hand/foot malformation in a fetus with an interstitial deletion of chromosome 2q that removes the SHFM5 locus. *Prenat Diagn* 2005; **25**: 39-44.
7 Van Buggenhout G, Van Ravenswaaij-Arts C, Mc Maas N *et al*: The del(2)(q32.2q33) deletion syndrome defined by clinical and molecular characterization of four patients. *Eur J Med Genet* 2005; **48**: 276-289.
8 Loscalzo ML, Galczynski RL, Hamosh A, Summar M, Chinsky JM, Thomas GH: Interstitial deletion of chromosome 2q32-34 associated with multiple congenital anomalies and a urea cycle defect (CPS I deficiency). *Am J Med Genet A* 2004; **128A**: 311-315.
9 Gallagher L, Becker K, Kearney G *et al*: Brief report: A case of autism associated with del(2)(q32.1q32.2) or (q32.2q32.3). *J Autism Dev Disord* 2003; **33**: 105-108.
10 Kondo K, Kaga K, Ogawa Y, Fukushima Y: Temporal bone histopathologic findings in a case of interstitial deletion of the long arm of chromosome 2 [del(2) (q31q33)]. *Int J Pediatr Otorhinolaryngol* 1999; **48**: 31-37.
11 Nixon J, Oldridge M, Wilkie AO, Smith K: Interstitial deletion of 2q associated with craniosynostosis, ocular coloboma, and limb abnormalities: cytogenetic and molecular investigation. *Am J Med Genet* 1997; **70**: 324-327.
12 Kreuz FR, Wittwer BH: Del(2q)--cause of the wrinkly skin syndrome? *Clin Genet* 1993; **43**: 132-138.
13 Hiroyama Y, Hatanaka H, Ikenoue T, Ishihara Y: Interstitial deletion of long arm of chromosome 2(q31q33). *Acta Paediatr Jpn* 1990; **32**: 563-565.
14 Palmer CG, Heerema N, Bull M: Deletions in chromosome 2 and fragile sites. *Am J Med Genet* 1990; **36**: 214-218.
15 Ramer JC, Mowrey PN, Robins DB, Ligato S, Towfighi J, Ladda RL: Five children with del (2)(q31q33) and one individual with dup (2)(q31q33) from a single family: review of brain, cardiac, and limb malformations. *Am J Med Genet* 1990; **37**: 392-400.
16 Glass IA, Swindlehurst CA, Aitken DA, McCrea W, Boyd E: Interstitial deletion of the long arm of chromosome 2 with normal levels of isocitrate dehydrogenase. *J Med Genet* 1989; **26**: 127-130.
17 Ramer JC, Ladda RL, Frankel CA, Beckford A: A review of phenotype-karyotype correlations in individuals with interstitial deletions of the long arm of chromosome 2. *Am J Med Genet* 1989; **32**: 359-363.
18 Dallapiccola B, Novelli G, Giannotti A: Deletion 2q31.3---2q33.3: gene dosage effect of ribulose 5-phosphate 3-epimerase. *Hum Genet* 1988; **79**: 92.
19 Miyazaki K, Yamanaka T, Ogasawara N: Interstitial deletion 2q32.1---q34 in a child with half normal activity of ribulose 5-phosphate 3-epimerase (RPE). *J Med Genet* 1988; **25**: 850-851.
20 Sumi S, Obayashi M, Murakami M *et al*: Interstitial deletion of the long arm of chromosome 2: a dose study on isocitrate dehydrogenase 1. *Jinrui Idengaku Zasshi* 1988; **33**: 461-467.
21 Benson K, Gordon M, Wassman ER, Tsi C: Interstitial deletion of the long arm of chromosome 2 in a malformed infant with karyotype 46,XX,del(2)(q31q33). *Am J Med Genet* 1986; **25**: 405-411.
22 Muneer RS, Payne-Howell RM, Altshuler G, Rennert OM: An interesting case of partial monosomy 2q and a 3q:12q translocation. *Pediatr Res* 1986; **20**: 269A.
23 Markovic S, Krstic M, Sulovic V, Radojkovic Z, Adzic S: Interstitial deletion of chromosome 2. *J Med Genet* 1985; **22**: 154-155.
24 Al-Awadi SA, Farag TI, Naguib K *et al*: Interstitial deletion of the long arm of chromosome 2: del(2)(q31q33). *J Med Genet* 1983; **20**: 464-465.
25 Buchanan PD, Rhodes RL, Stevenson CE, Jr.: Interstitial deletion 2q31 leads to q33. *Am J Med Genet* 1983; **15**: 121-126.
26 Franceschini P, Cirillo Silengo M, Davi G, Bianco R, Biagioli M: Interstitial deletion of the long arm of chromosome 2 (q31q33) in a girl with multiple anomalies and mental retardation. *Hum Genet* 1983; **64**: 98.
27 Pai GS, Rogers JF, Sommer A: Identical multiple congenital anomalies/mental retardation (MCA/MR) syndrome due to del(2)(q32) in two sisters with intrachromosomal insertional translocation in their father. *Am J Med Genet* 1983; **14**: 189-195.
28 Young RS, Shapiro SD, Hansen KL, Hine LK, Rainosek DE, Guerra FA: Deletion 2q: two new cases with karyotypes 46,XY,del(2)(q31q33) and 46,XX,del(2)(q36). *J Med Genet* 1983; **20**: 199-202.
29 Shabtai F, Klar D, Halbrecht I: Partial monosomy of chromosome 2. Delineable syndrome of deletion 2 (q23-q31). *Ann Genet* 1982; **25**: 156-158.
30 Taysi K, Dengler DR, Jones LA, Heersma JR: Interstitial deletion of the long arm of chromosome 2: case report and review of literature. *Ann Genet* 1981; **24**: 245-247.




**Supplementary Figure S2** Result of MLPA analyses in patient 53B. The P155 kit contains 11 probes for *COL3A1* (fragments 4, 9, 12, 15, 19, 21, 31, 35, 38, 40, and 44; red bars) and 14 control MLPA probes located on chromosomes 1, 5, 6, 7, 9, 11, 13, 17, 18, and 22 (fragments 1, 6, 7, 10, 14, 18, 20, 24, 27, 30, 36, 39, 42, and 45) as well as 14 probes for *TNXB*, two for *CYP21A1P*, and one probe for each of the genes *C4B*, *LTA*, *BAK1*, and *CREBL1*. The normal range of ±30% of the normalized relative peak areas is given by dotted lines. Note that in patient 53B all probes for *COL3A1* showed a reduced relative peak area outside the normal range (for reduced peaks, the corresponding exons are denoted) and that there are two MLPA probes for *COL3A1* exon 1.

5

## Contribution of Authors

| | |
|---|---|
| **Janine Meienberg** | **MLPA analysis, array and breakpoint analyses, writing of the manuscript** |
| Marianne Rohrbach | Clinical examination, writing of the manuscript |
| Stefan Neuenschwander | Analysis of array data, editing of the manuscript |
| Katharina Spanaus | Laboratory blood analyses, editing of the manuscript |
| Cecilia Giunta | Collagen analyses, editing of the manuscript |
| Sira Alonso | MLPA analysis, editing of the manuscript |
| Eliane Arnold | Sequence analyses, editing of the manuscript |
| Caroline Henggeler | Sequence analyses, MLPA analysis, editing of the manuscript |
| Stephan Regenass | Laboratory blood analyses, editing of the manuscript |
| Andrea Patrignani | Array analysis, editing of the manuscript |
| Silvia Azzarello-Burri | Clinical examination, editing of the manuscript |
| Bernhard Steiner | Clinical examination, editing of the manuscript |
| Anders OH Nygren | Development of MLPA kit P155, editing of the manuscript |
| Thierry Carrel | Clinical examination, editing of the manuscript |
| Beat Steinmann | Clinical examination, editing of the manuscript |
| Gabor Matyas | Conceiving and planning the study, writing and editing of the manuscript |

# Appendix 2 Supplementary Information to «New Insights into the Performance of Human Whole-Exome Capture Platforms» (2.1.2)

Nucleic Acids Research - Methods

**SUPPLEMENTARY INFORMATION**

# New insights into the performance of human whole-exome capture platforms

Janine Meienberg[1,#], Katja Zerjavic[1,#], Irene Keller[2], Michal Okoniewski[3,4], Andrea Patrignani[3], Katja Ludin[5], Zhenyu Xu[6], Beat Steinmann[7], Thierry Carrel[8], Benno Röthlisberger[5], Ralph Schlapbach[3], Rémy Bruggmann[9], and Gabor Matyas[1,8,10,*]

[1] Center for Cardiovascular Genetics and Gene Diagnostics, Foundation for People with Rare Diseases, Schlieren-Zurich, CH-8952, Switzerland
[2] Department of Clinical Research, University of Berne, Berne, CH-3010, Switzerland
[3] Functional Genomics Center Zurich, Zurich, CH-8057, Switzerland
[4] Division of Scientific IT Services, ETH Zurich, Zurich, CH-8092, Switzerland
[5] Division of Medical Genetics, Center for Laboratory Medicine, Aarau, CH-5001, Switzerland
[6] Sophia Genetics SA, Lausanne, CH-1015, Switzerland
[7] Division of Metabolism, University Children's Hospital, Zurich, CH-8032, Switzerland
[8] Department of Cardiovascular Surgery, University Hospital, Berne, CH-3010, Switzerland
[9] Interfaculty Bioinformatics Unit and Swiss Institute of Bioinformatics, University of Berne, Berne, CH-3012, Switzerland
[10] Zurich Center of Integrative Human Physiology, University of Zurich, Zurich, CH-8057, Switzerland

* To whom correspondence should be addressed. Tel: +41 43 433 86 86; Fax: +41 43 433 86 85; Email: matyas@genetikzentrum.ch

[#] The authors wish it to be known that, in their opinion, the first 2 authors should be regarded as joint First Authors.

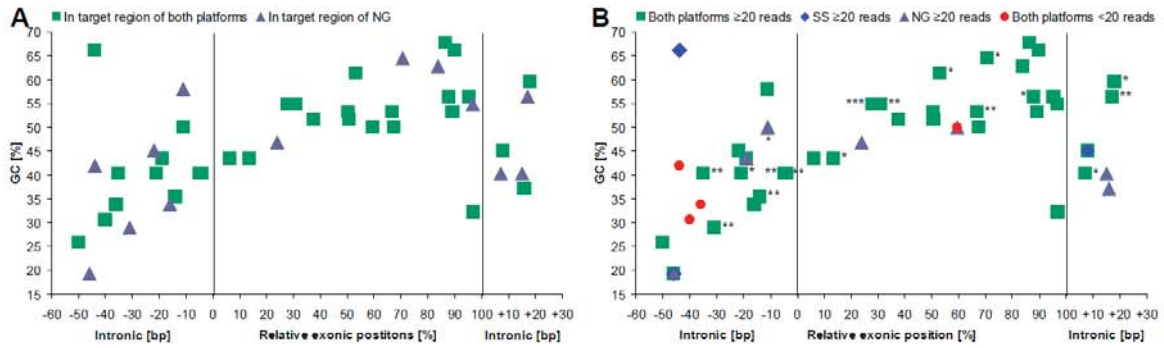



**Figure S1.** Preliminary study: Positions of selected heterozygous sequence variants detected by Sanger sequencing in our region of interest (exons with -50-bp and +20-bp flanking intronic sequences) as a function of the GC content of 30-bp flanking sequences. (**A**) Occurrence of variants in the designed target region of each platform. (**B**) Platforms providing ≥20 reads at heterozygous variant positions. Note that, contrary to expectations, not all exonic positions are within the designed target region of the Agilent platform. Furthermore, none of the two platforms achieved complete coverage of all exonic positions at ≥20 reads. Exonic positions are given in percentage [%] relative to the length of the corresponding exon, whereas intronic positions are given as absolute positions in base pairs [bp]. SS, Agilent SureSelect v4+UTR; NG, NimbleGen SeqCap v3; *, heterozygous in one additional DNA sample with the same symbol; **, heterozygous in two additional DNA samples with the same symbol; ***, heterozygous in three additional DNA samples with the same symbol.

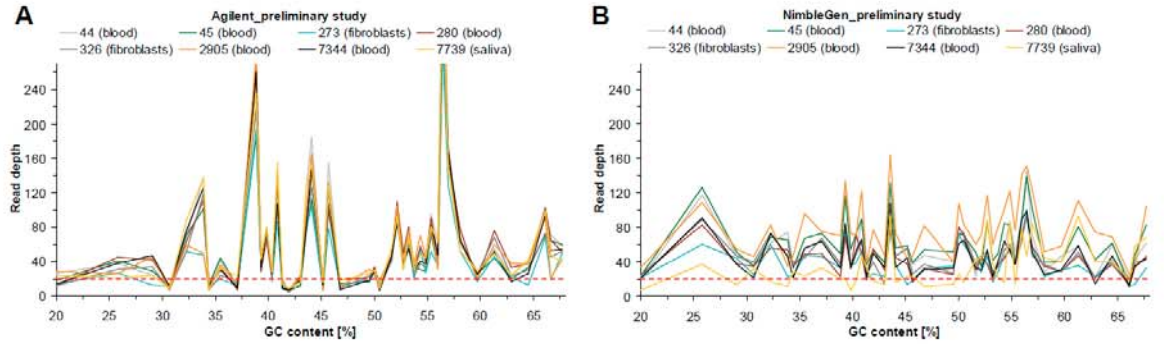
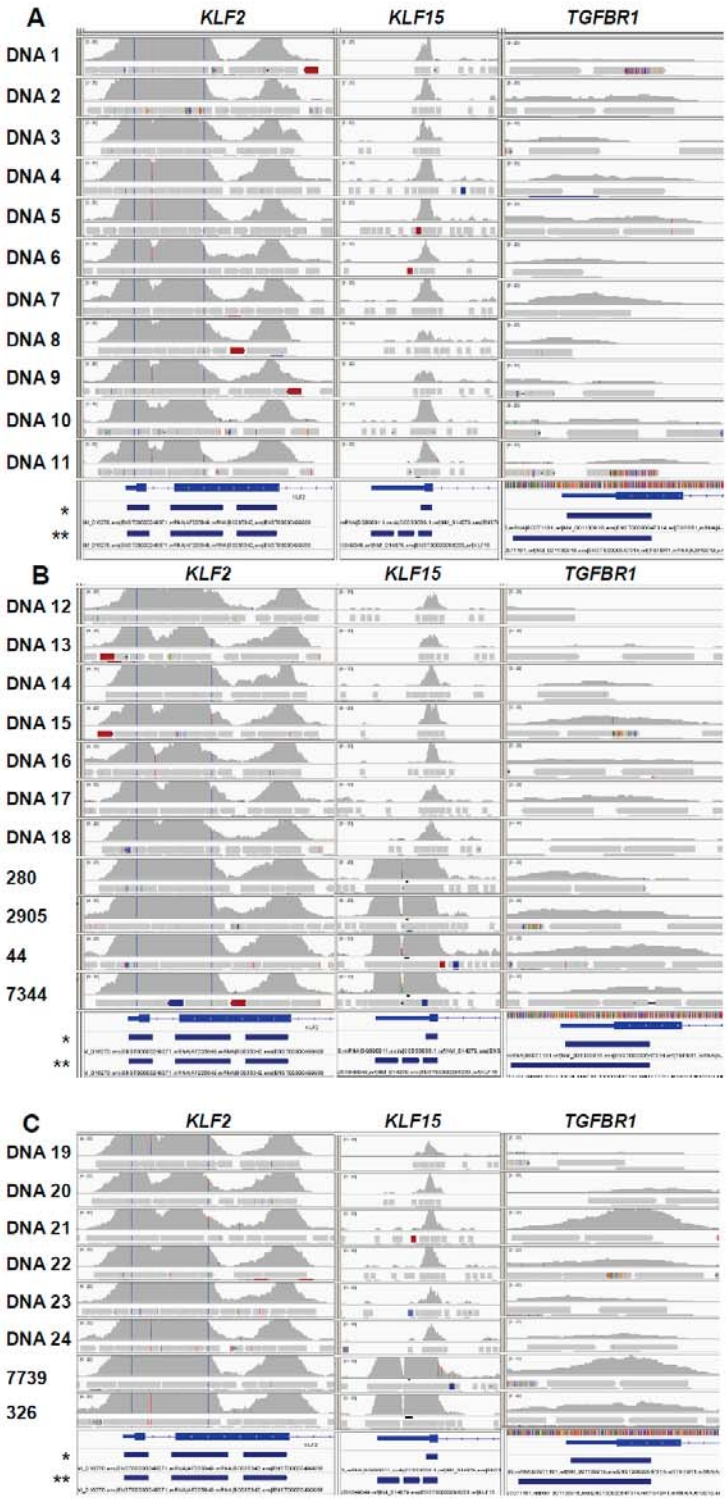



**Figure S2.** Preliminary study: Read depth for 44 different SNV positions within our region of interest (coding exons and -50-bp and +20-bp flanking intronic sequences) called as heterozygous in at least one of the eight samples (the six samples from this study as well as 45 and 273) plotted against the GC content of 30-bp flanking sequences. (**A**) Agilent (SureSelect Human All Exon kit v4+UTR). (**B**) NimbleGen (SeqCap EZ Exome v3). Red dashed line indicates our limit of 20 reads. Note the high number of positions failed to reach the limit of 20 reads by the Agilent platform.




Although DNA samples extracted from blood performed slightly better, no distinct performance difference among the 24 additional DNA samples sequenced at 60× using Agilent was observed in selected GC-rich exons of *KLF2*, *KLF15*, and *TGFBR1*, confirming our observations on the six samples sequenced at 100× using Agilent. Unexpectedly, for one extremely GC-rich (83%) exon, which was not covered sufficiently even when sequenced at 100×, one DNA extracted from saliva and sequenced at 60× achieved a mean of 18 reads (*TGFBR1* in DNA 21).

**Figure S3.** Comparison of enrichment efficiency for GC-rich exons in the six samples of the present study and 24 additional DNA samples. Integrative Genomics Viewer (IGV) print screens showing coverage track (0-20 reads) for *KLF2* exons 1 and 2, *KLF15* exon 3, and *TGFBR1* exon 1, which have high GC content. (**A**, **B**) DNA samples extracted from blood. (**C**) DNA samples extracted from saliva (DNA 19-24, 7739) or fibroblasts (326). For the six samples of this study (44, 280, 326, 2905, 7344, 7739), WES data of V2 using Agilent SureSelect v5+UTR at 100× are shown. The additional 24 samples (DNA 1-24) were sequenced by V2 using Agilent SureSelect v5 at 60×. *, designed target region of Agilent SureSelect v5; **, designed target region of Agilent SureSelect v5+UTR.

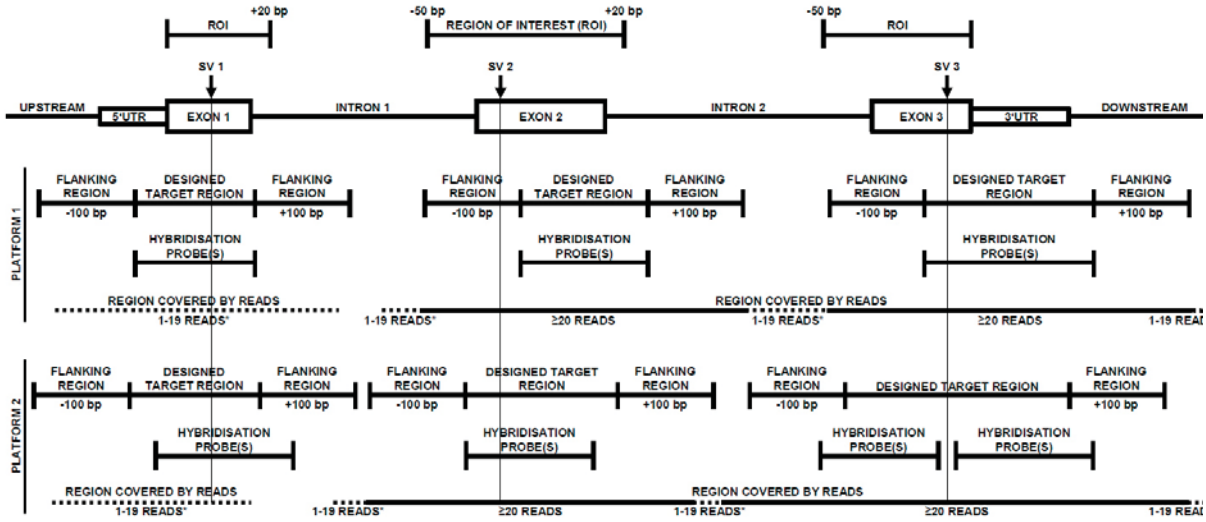



**Figure S4.** Schematic presentation of terms used in this study. SV, sequence variant; *insufficient coverage e.g. due to GC-rich regions or distance to hybridisation probes; closed lines, regions with defined length and positions; open lines, regions with variable length and positions. Note that according to Agilent's definition the designed target region is completely covered by hybridisation probes (cf. platform 1), whereas NimbleGen and Illumina define it as the region intended/designed to be enriched (cf. platform 2).

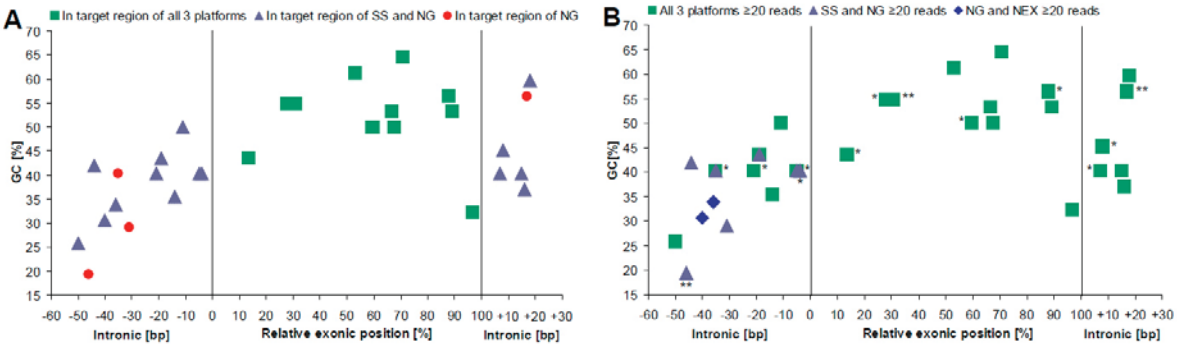



**Figure S5.** Positions of selected heterozygous sequence variants detected by Sanger sequencing in a region of interest most relevant for mutation screening (exons with -50-bp and +20-bp flanking intronic sequences) as a function of the GC content of 30-bp flanking sequences. (**A**) Occurrence of variants in the designed target region of each platform. (**B**) Platforms providing ≥20 reads at heterozygous variant positions (results per vendor are shown in Supplementary Figure S6). Note that all exonic positions are in designed target regions and completely covered by all three enrichment platforms with ≥20 reads and that Agilent and particularly Illumina covered variants outside its designed target region with ≥20 reads. Exonic positions are given in percentages [%] relative to the length of the corresponding exon, whereas intronic positions are given as absolute positions in base pairs [bp]. SS, Agilent SureSelect v5+UTR; NG, NimbleGen SeqCap v3+UTR; NEX, Illumina Nextera Expanded Exome; *, heterozygous in one additional DNA sample with the same symbol; **, heterozygous in two additional DNA samples with the same symbol.

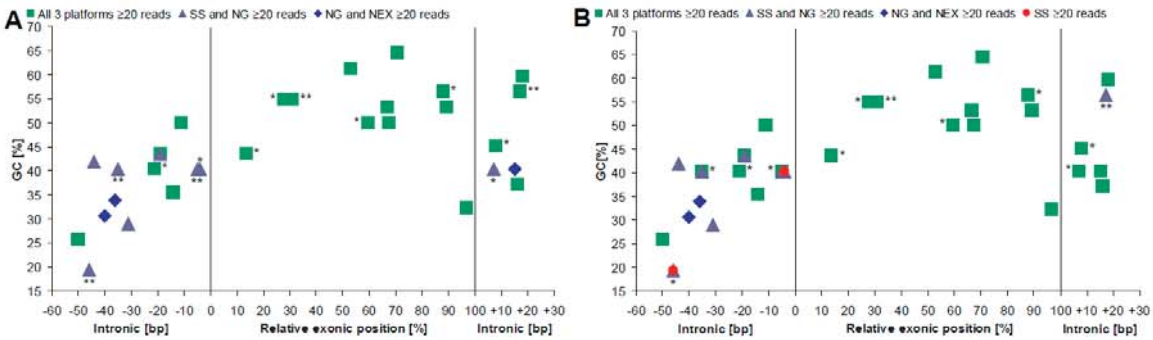



**Figure S6.** Enrichment performance per vendor for selected heterozygous sequence variants detected by Sanger sequencing in our region of interest (exons with -50-bp and +20-bp flanking intronic sequences) as a function of the GC content of 30-bp flanking sequences. (**A**) Platforms performed by the same vendor (V1) providing ≥20 reads at heterozygous variant positions. (**B**) Platforms performed by different vendors (V2-V4) providing ≥20 reads at heterozygous variant positions. Note that all exonic positions are completely covered by all three enrichment platforms with ≥20 reads (regardless of vendor). Exonic positions are given in percentage [%] relative to the length of the corresponding exon, whereas intronic positions are given as absolute positions in base pairs [bp]. SS, Agilent SureSelect v5+UTR; NG, NimbleGen SeqCap v3+UTR; NEX, Illumina Nextera Expanded Exome; *, heterozygous in one additional DNA sample with the same symbol; **, heterozygous in two additional DNA samples with the same symbol.

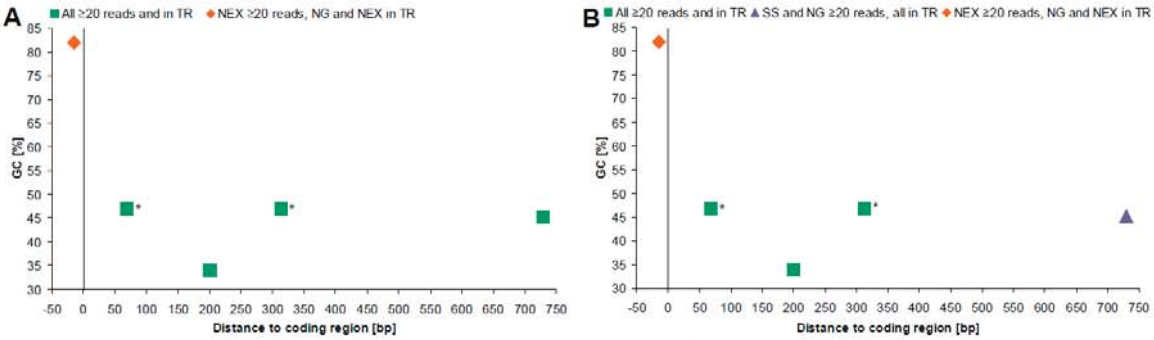



**Figure S7.** Enrichment performance per vendor for DNA samples with selected heterozygous sequence variants detected by Sanger sequencing in UTR as a function of the GC content of 30-bp flanking sequences. (**A**) Results of the same vendor for all platforms (V1). (**B**) Data of different vendors (V2-V4). Positions of variants are given as distance to coding region in base pairs [bp] (i.e. -50 to -1 for 5'UTR and 1 to 750 for 3'UTR). TR, designed target region; SS, Agilent SureSelect v5+UTR; NG, NimbleGen SeqCap v3+UTR; NEX, Illumina Nextera Expanded Exome; *, heterozygous in one additional DNA sample with the same symbol. Note that one variant position in a GC-rich (82%) exon (orange) is located within the designed target region of both NimbleGen (NG) and Illumina (NEX) but is only covered by Illumina (NEX) at 20× (regardless of vendor).

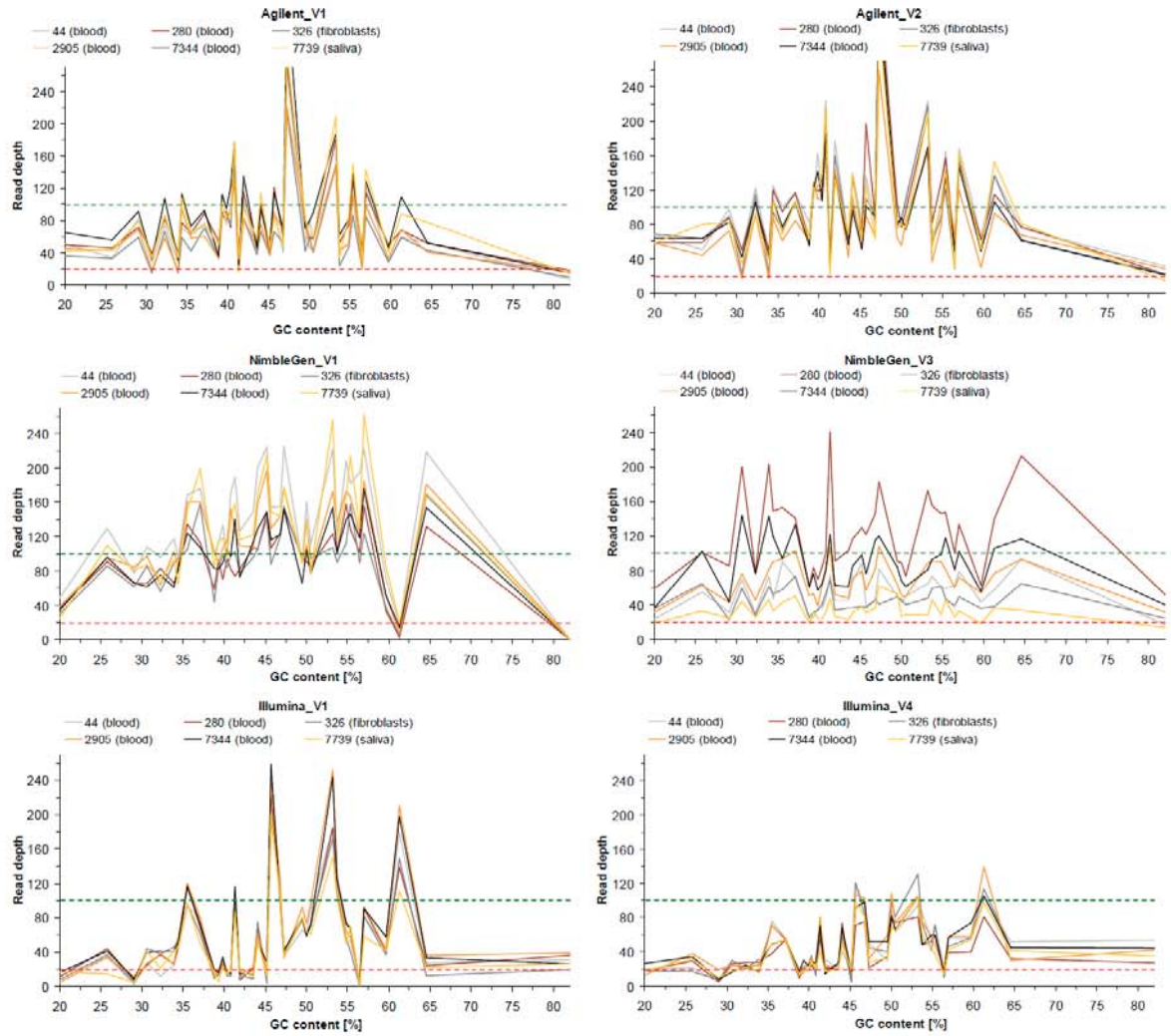



**Figure S8.** Read depth of all six DNA samples for 30 different SNV positions within our region of interest (coding exons and -50-bp and +20-bp flanking intronic sequences) and five in UTR plotted against the GC content of 30-bp flanking sequences. Green dashed lines indicate expected read depth of 100 (reads) and red dashed lines denote read depth of 20 (reads). Note the performance of Agilent and NimbleGen compared to Illumina as well as the intersample variation in the performance of V3 using the NimbleGen platform.

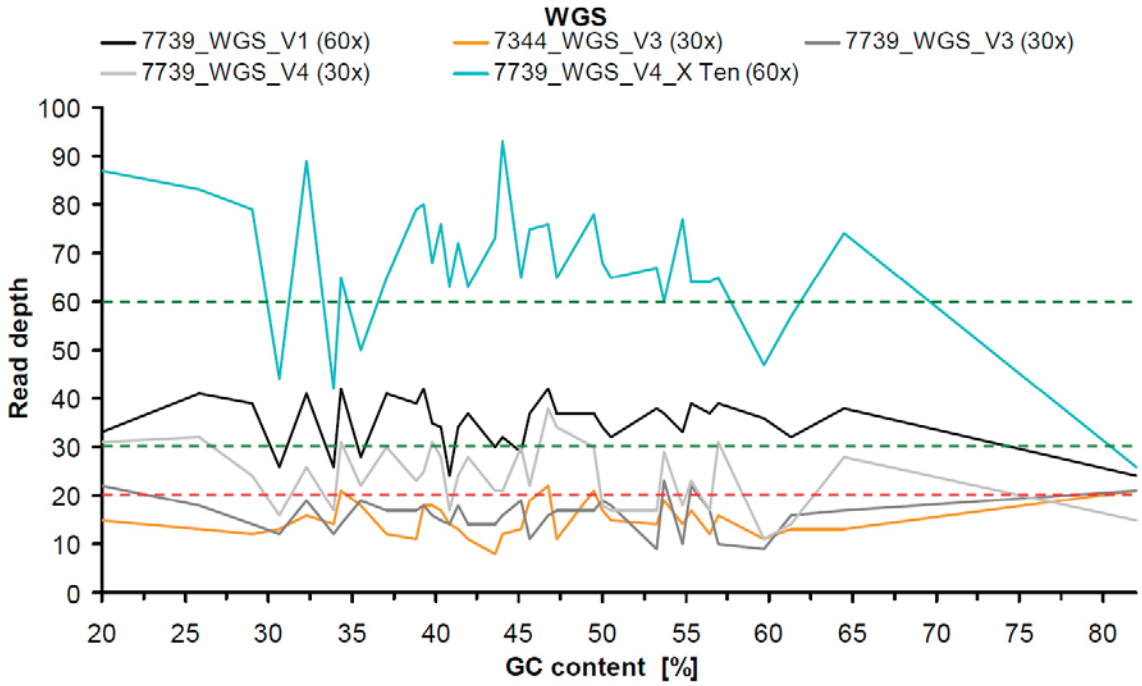



**Figure S9.** Read depth of all five WGS data sets for 30 different SNV positions within our region of interest (coding exons and -50-bp and +20-bp flanking intronic sequences) and five in UTR plotted against the GC content of 30-bp flanking sequences. Green dashed lines indicate expected read depth of 60 and 30 (reads), respectively, and red dashed line denotes read depth of 20 (reads). Note that at 30× V4 performed better than V3 and that only WGS data at 60× (V1 and V4_X Ten) reached for all positions the limit of 20 reads (cf. Supplementary Figure S17). X Ten, HiSeq X Ten system.

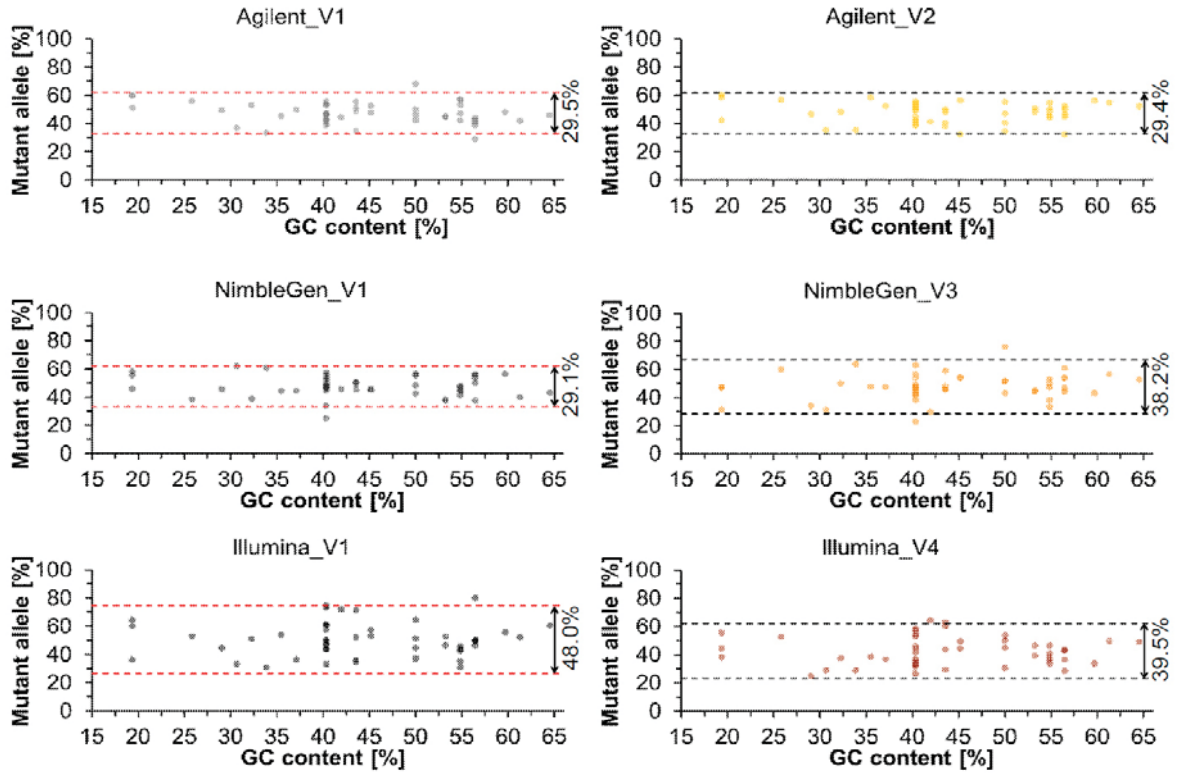



**Figure S10.** Percentage of non-reference (mutant) alleles called for heterozygous variants detected by Sanger sequencing in our region of interest (exons with -50-bp and +20-bp flanking intronic sequences) displayed for all six platform-vendor combinations. 30 different heterozygous SNVs listed according to GC content of 30-bp flanking sequences. Shown are values of all six DNA samples. Dashed lines indicate an interval within which 95% of the percentage values of non-reference alleles lie (calculated according to the Student's ***t*** distribution as the mean of ***n*** percentage values ± critical ***t*** value ($t_{crit,n-1}$) × ***SD*** using $n = 49$, $t_{crit} = 2.011$).

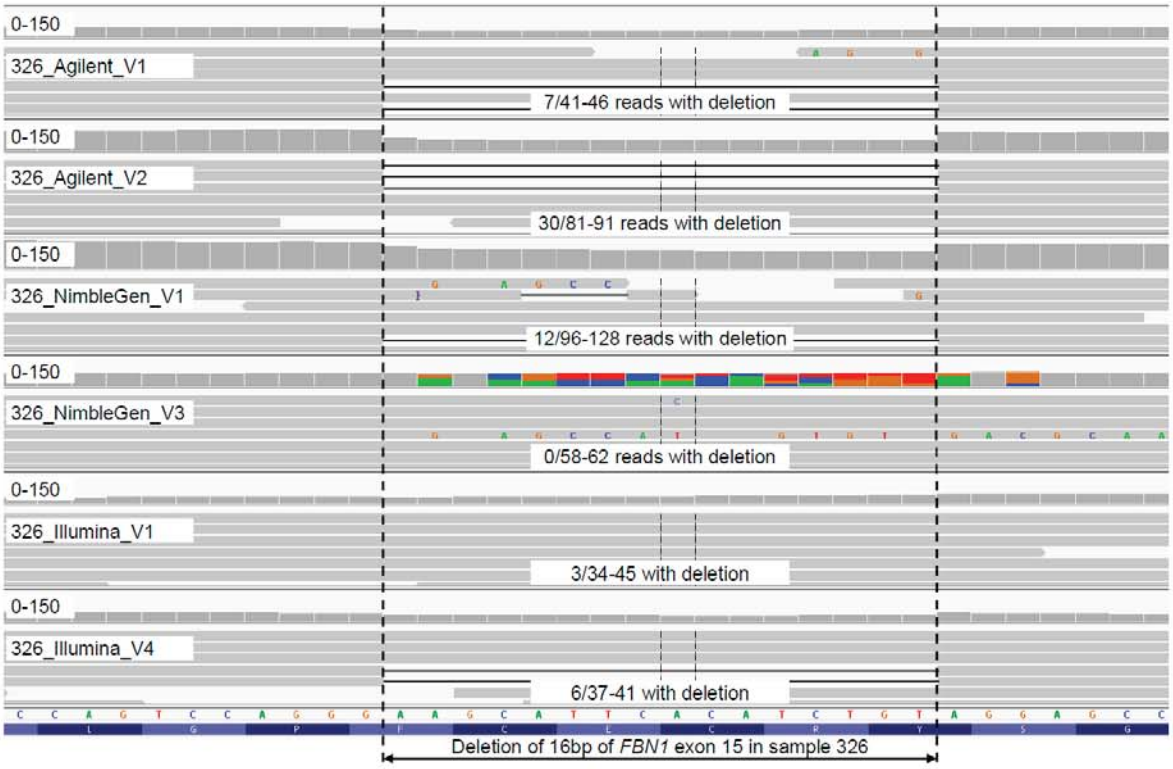



**Figure S11**. Aligned reads for a heterozygous exonic deletion of 16 bp. Sections of six reads per data set covering the deleted region in sample 326 are presented using the Integrative Genomics Viewer (IGV). Gray bars/arrows denote aligned reads, letters indicate mismatched bases, and black horizontal lines represent deletions. Display range for read depth is set to 0-150 and counts only called bases (no deletions). Coloured bars indicate called mismatches. Note that in the data of V3 using NimbleGen the deleted alleles are called as a series/consecutive of mismatches rather than a deletion, most likely due to the bioinformatics data analysis workflow of V3 (e.g., improper realignment of the region) and that V2 using Agilent called the largest fraction of reads with this deletion.

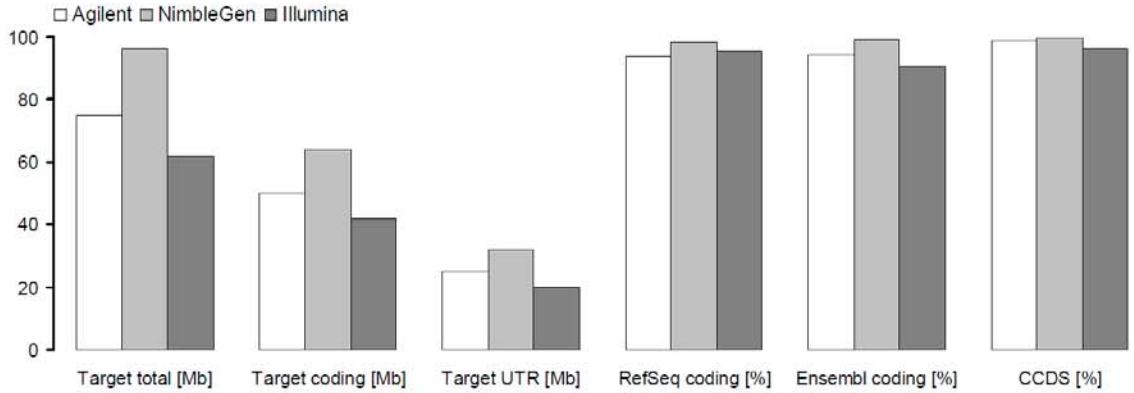



**Figure S12.** Designed target regions of the three updated exome enrichment platforms. Total number of bases within the designed target region (target total), its distribution on coding (target coding) and untranslated regions (target UTR), and coverage of coding exons of the gene databases RefSeq, Ensembl, and CCDS are presented for Agilent (SureSelect v5+UTR), NimbleGen (SeqCap v3+UTR), and Illumina (Nextera Expanded Exome) according to the companies' specifications (Supplementary Table S2). CCDS, Consensus Coding Sequences.

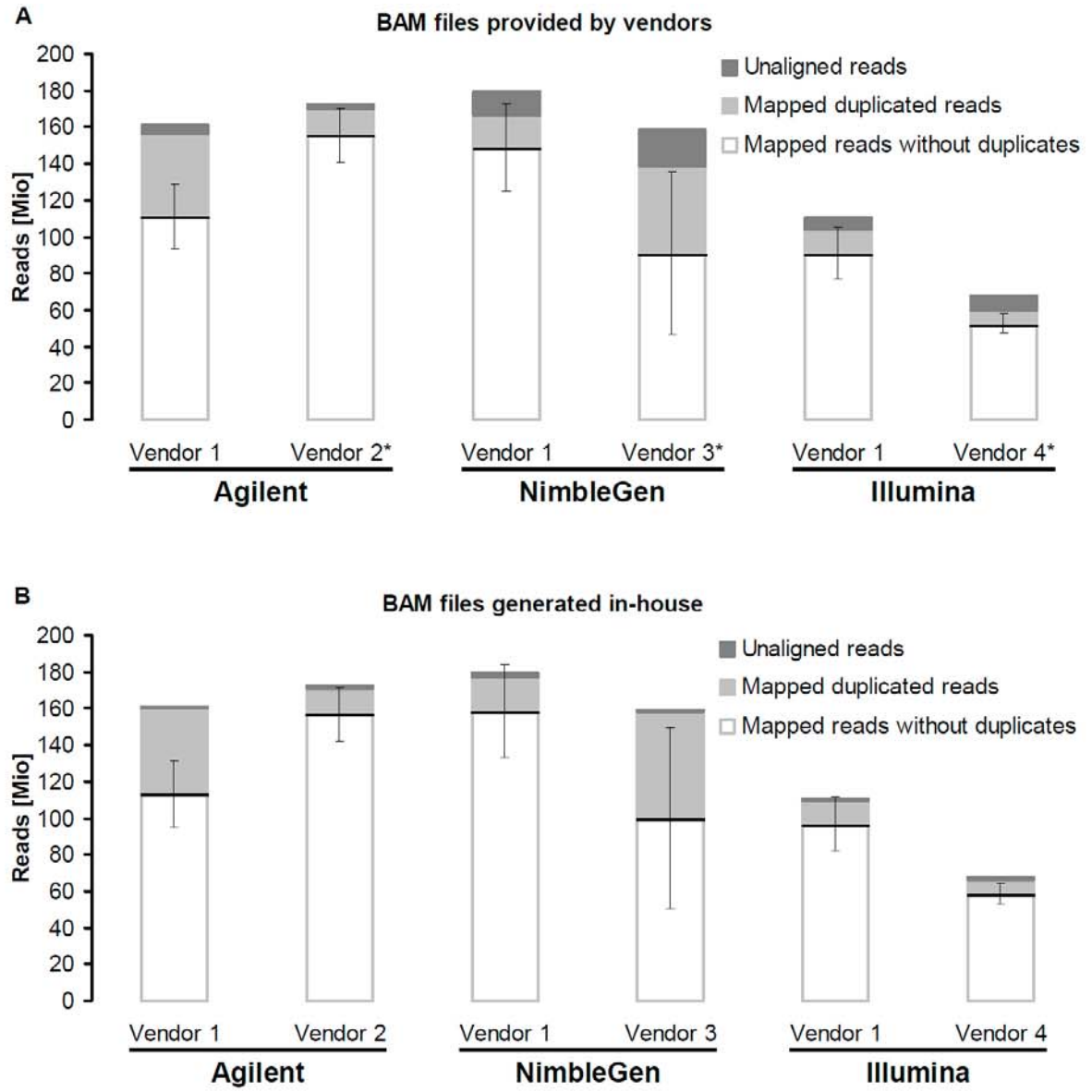



**Figure S13.** Reads generated using the three different enrichment platforms (Agilent, NimbleGen, and Illumina) applied by different vendors. (**A**, **B**) Number of unaligned reads, mapped duplicated reads, and remaining reads used for analysis (mapped reads without duplicates) according to BAM files provided by vendors (**A**) and in-house generated BAM files using the same bioinformatics pipeline (**B**). *Note that bioinformatics workflow is different among vendors and that vendor 3 and vendor 4 provided unique mapped reads in contrast to total mapped reads (cf. Supplementary Table S3). Given are means of all six DNA samples (n=6); error bars indicate 95% confidence intervals.

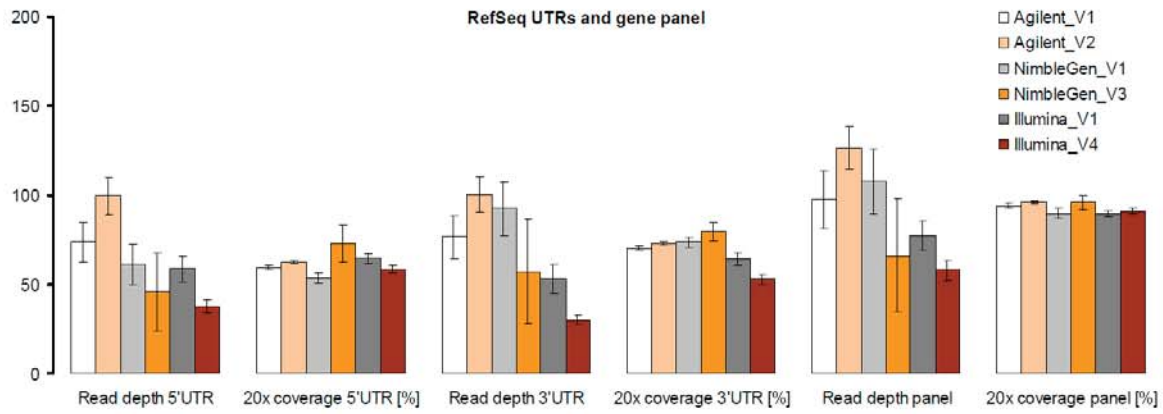



**Figure S14.** Enrichment efficiency of the three updated exome enrichment platforms (Agilent, NimbleGen, and Illumina) performed by different vendors (V1, V2, V3, and V4) for UTR of RefSeq exons and a panel of eight genes (cf. Supplementary Figure S27 for data on a larger set of clinically relevant exons). Mean read depth and percentage of coverage at 20× for 5'UTR and 3'UTR of the RefSeq database as well as for exons (coding and UTR) of our panel of eight genes (panel) specified in Supplementary Table S1. Given are means of all six DNA samples; error bars indicate 95% confidence intervals. Values were calculated using the SeqMonk program (www.bioinformatics.babraham.ac.uk/projects/seqmonk) and are presented in Supplementary Tables S6, S12, S14, and S15.

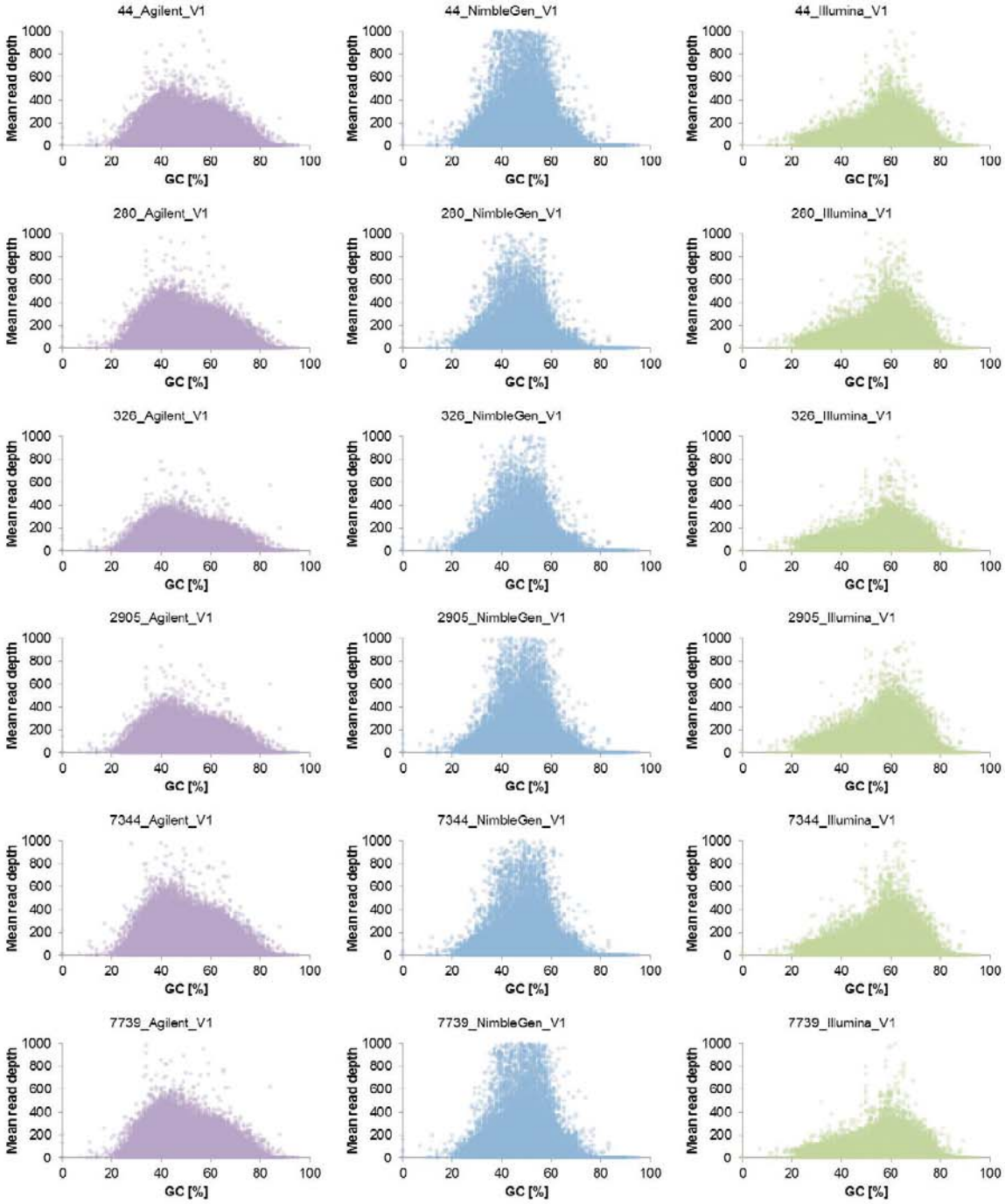



**Figure S15.** Enrichment bias in terms of read depth owing to GC content for the three enrichment platforms (Agilent, NimbleGen, Illumina) performed by the same vendor (V1). Scatter plots showing GC content and achieved read depth of RefSeq exons (coding and UTR).

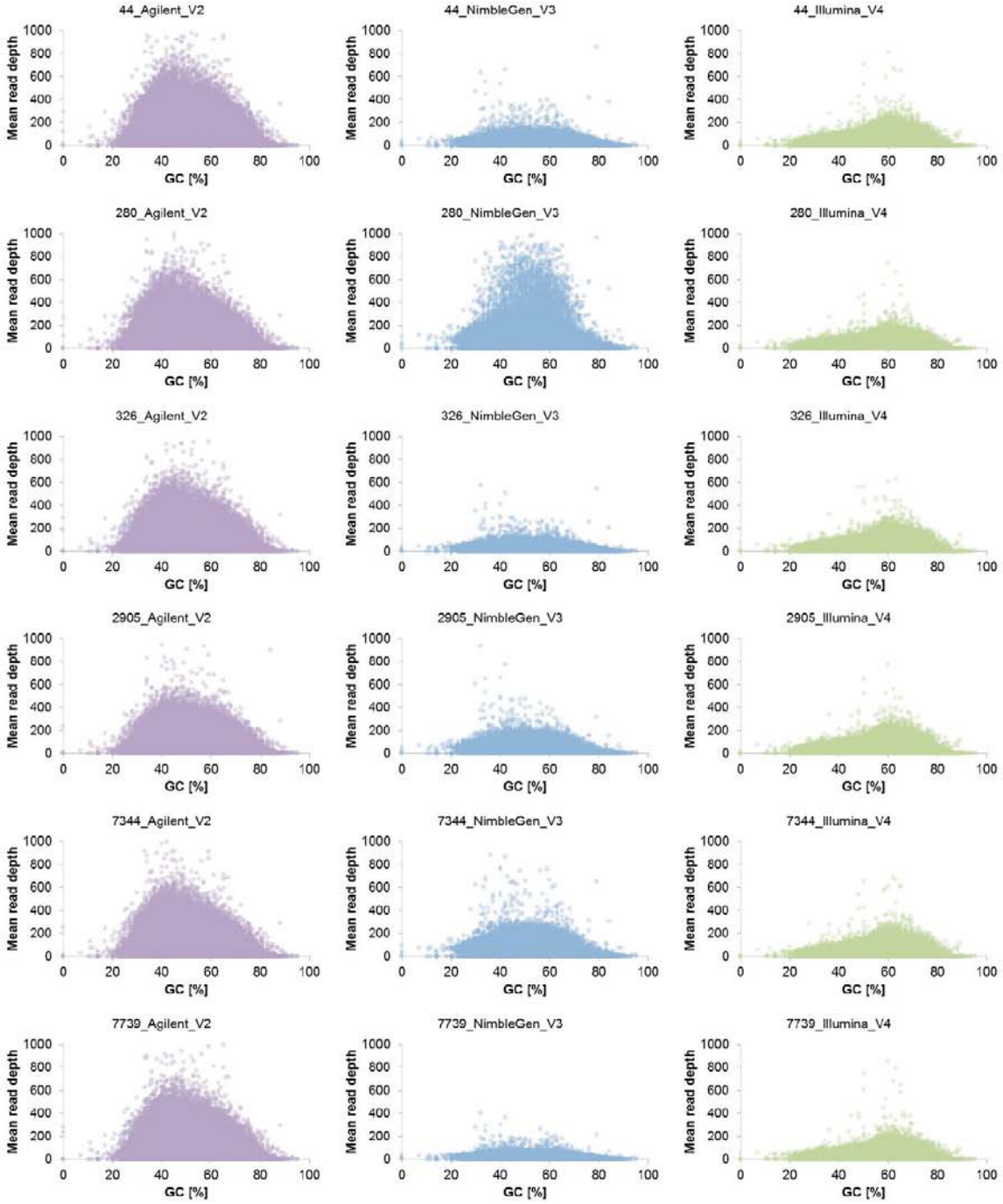



**Figure S16.** Enrichment bias in terms of read depth owing to GC content for the three enrichment platforms (Agilent, NimbleGen, Illumina) performed by different vendors (V2-V4). Scatter plots showing GC content and achieved read depth of RefSeq exons (coding and UTR).

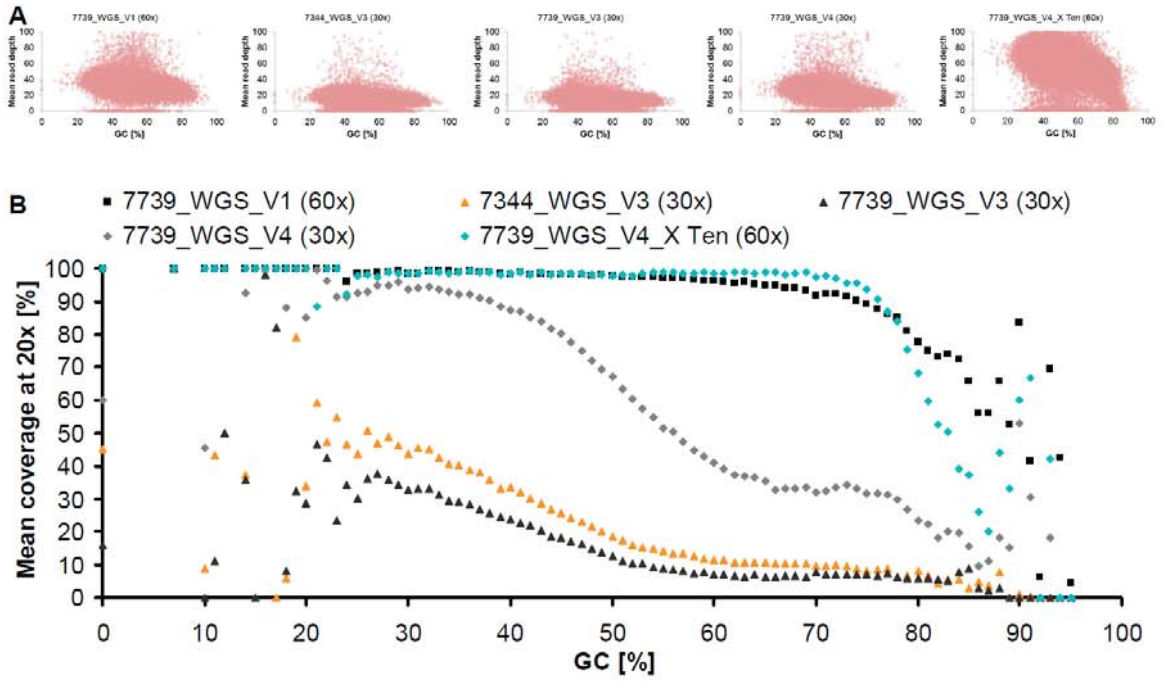



**Figure S17.** Influence of GC content on the mean read depth and coverage of RefSeq exons in whole genome sequencing (WGS). (**A**) Scatter plots showing GC content and achieved mean read depth. (**B**) Mean coverage at 20× per GC content. Data are shown for RefSeq exons (coding and UTR) for all five WGS data sets, i.e. for 7739 sequenced by V1 at 60× and by V3 and V4 at 30× as well as for 7344 sequenced by V3 at 30× on HiSeq2000/2500 using Illumina's TruSeq DNA PCR-Free Sample Preparation Kit and for 7739 sequenced by V4 at 60× on a HiSeq X Ten system using Illumina's TruSeq Nano DNA Sample Preparation Kit. Note: GC-rich exons were better covered by WGS than by WES, thereby WGS performed by V3 resulted in comparable coverage for both samples in spite of different DNA sources, whereas V4 showed better WGS performance at 30× than V3 and, as expected, WGS at 60× was superior to sequencing at 30× (cf. PCR-free sample preparation (V1) performed better than TruSeq Nano (V4_X Ten), especially above 80% GC content).

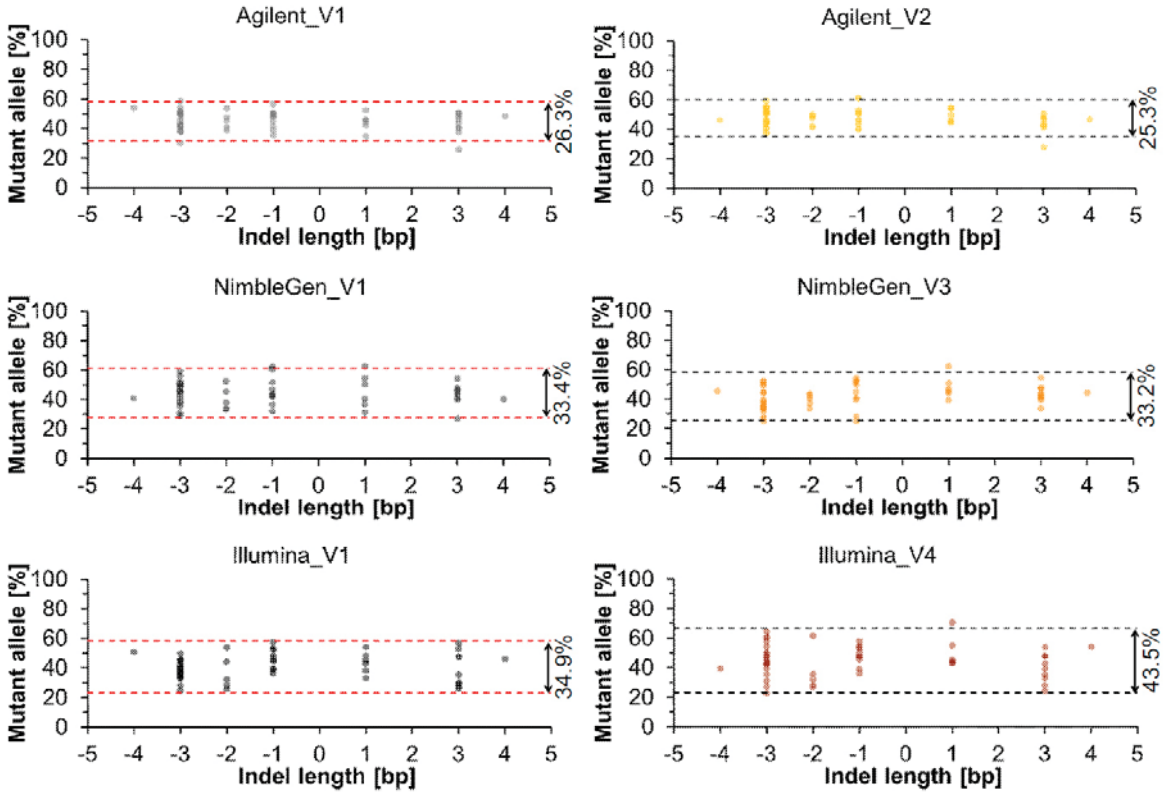



**Figure S18.** Relative proportion of non-reference (mutant) alleles for indels called in the VCF files provided by vendors (V1-V4). The analysis was restricted to shared heterozygous variants within the designed target regions of the three platforms (Agilent, NimbleGen, and Illumina) located in exons completely (100%) covered at 20× by all six platform-vendor combinations. Heterozygous indels listed according to their length, thereby negative and positive values indicate deletions and insertions, respectively. Shown are values of all six DNA samples. Dashed lines indicate an interval within which 95% of the percentage values of non-reference alleles lie (calculated according to the Student's *t* distribution as the mean of *n* percentage values ± critical *t* value ($t_{crit,n-1}$) × *SD* using $n$ = 51, $t_{crit}$ = 2.009).

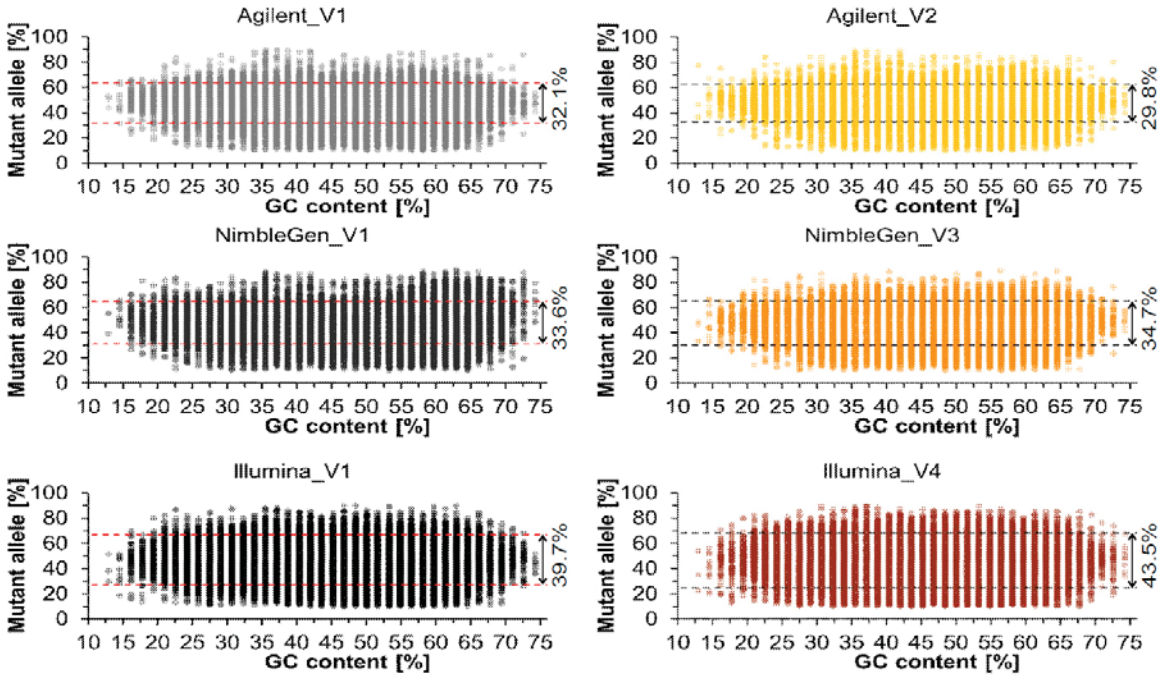



**Figure S19.** Percentages of non-reference (mutant) alleles for heterozygous SNVs called in our gVCF files generated by the same in-house bioinformatics pipeline for all six platform-vendor combinations. Fraction of biallelic non-reference alleles for shared heterozygous variants within the platforms' designed target region and 50-bp flanking sequences achieving ≥20 reads and >30 quality scores by all six platform-vendor combinations are displayed according to the GC content of 30-bp flanking sequences. Shown are values of all six DNA samples. Variant positions with more than one different non-reference allele (non-biallelic) as well as variant calls with alternative allele percentages outside 10-90% were excluded. Dashed lines indicate an interval within which 95% of the percentage values of non-reference alleles lie (calculated according to the Student's *t* distribution as the mean of *n* percentage values ± critical *t* value ($t_{crit,n-1}$) × *SD* using $n$ = 92'158, $t_{crit}$ = 1.960).

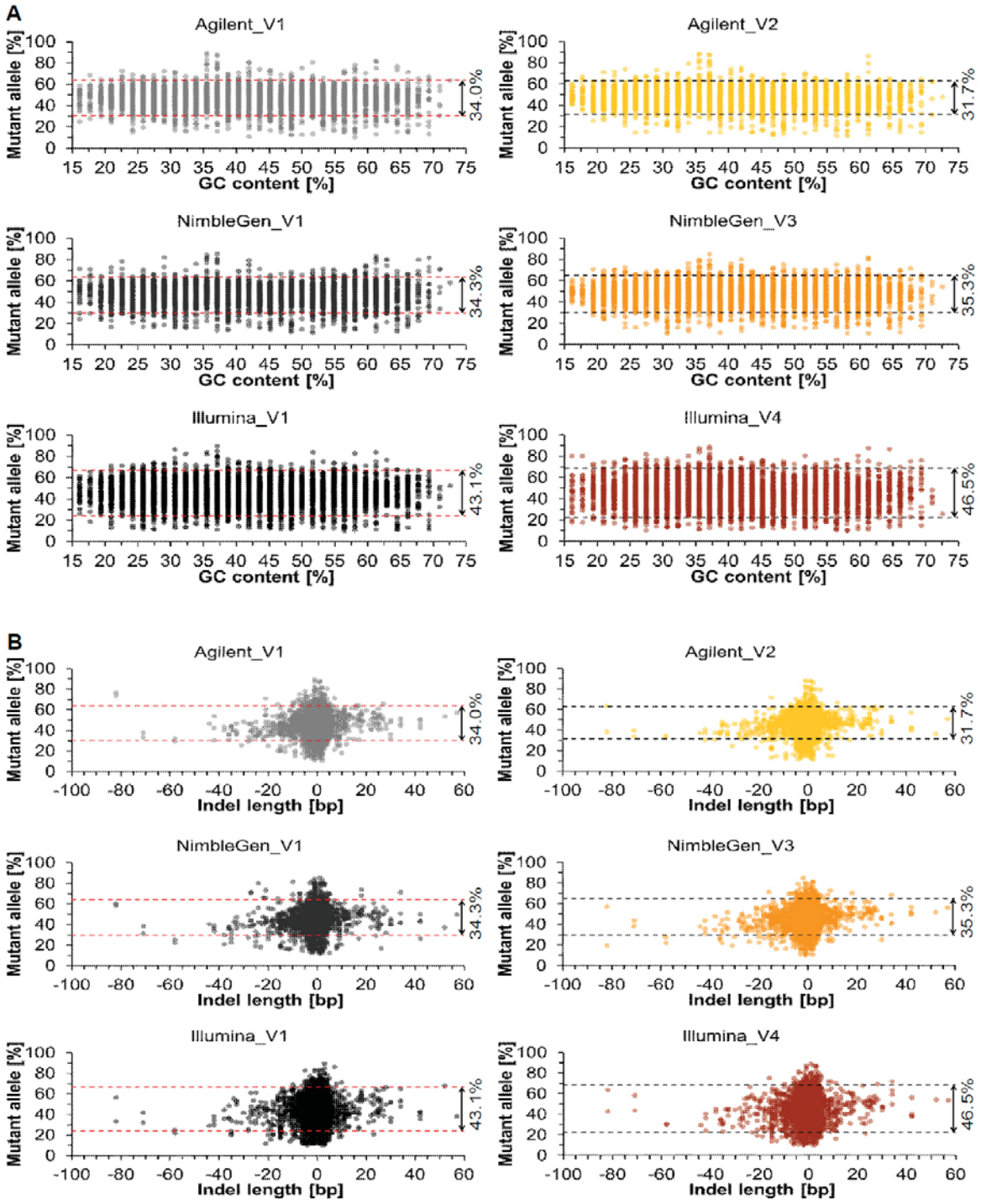



**Figure S20.** Percentages of non-reference (mutant) alleles for heterozygous indels called in our gVCF files generated by the same in-house bioinformatics pipeline for all six platform-vendor combinations. Fraction of biallelic non-reference alleles for shared heterozygous variants within the platforms' designed target region and 50-bp flanking sequences achieving ≥20 reads and >30 quality scores by all six platform-vendor combinations are displayed. (**A**) Diagrammed according to the GC content of 30-bp flanking sequences. (**B**) Diagrammed according to their length, thereby negative and positive values indicate deletions and insertions, respectively. Shown are values of all six DNA samples. Variant positions with more than one different non-reference allele (non-biallelic) as well as variant calls with alternative allele percentages outside 10-90% were excluded. Dashed lines indicate an interval within which 95% of the percentage values of non-reference alleles lie (calculated according to the Student's *t* distribution as the mean of *n* percentage values ± critical *t* value ($t_{crit,n-1}$) × *SD* using *n* = 5'645, $t_{crit}$ = 1.960).

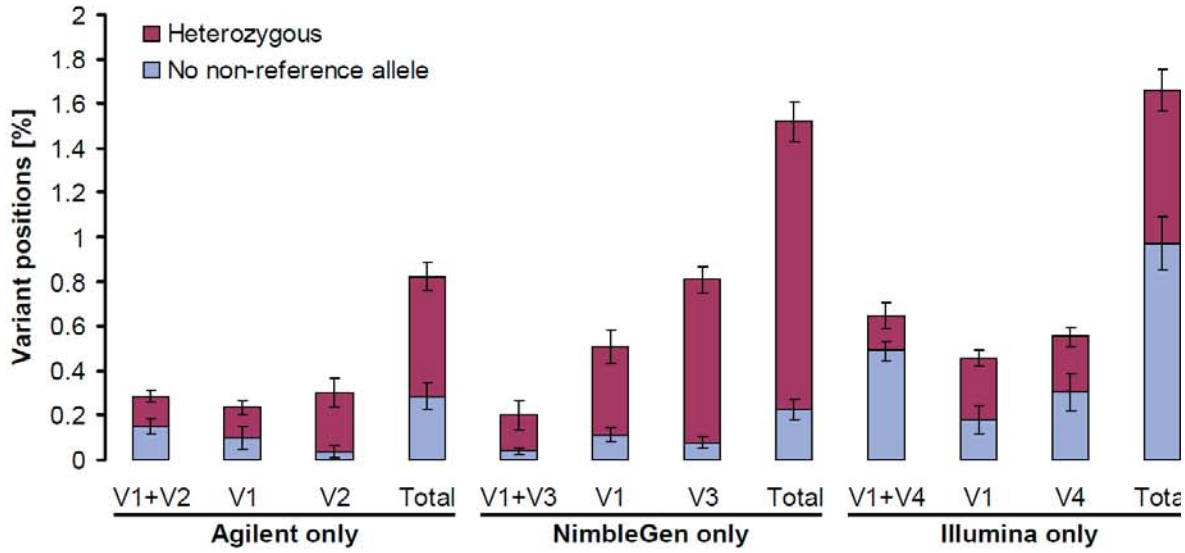



**Figure S21.** Differences in variant calling among platforms (Agilent, NimbleGen, Illumina) and vendors (V1-V4) in our gVCF files generated by the same in-house bioinformatics pipeline. Displayed are the relative proportions of heterozygous variant positions either called (heterozygous) or missed (no non-reference allele) by only one of the three platforms (i.e. called as heterozygous by only one or by all but one platform). Genomic positions within the platforms' designed target regions and 50-bp flanking sequences which achieved ≥20 reads and >30 quality scores for all six platform-vendor combinations were considered for this analysis (cf. Supplementary Table S3), thereby excluding variant positions with more than one different alternative allele (non-biallelic) or heterozygous calls with allele fractions outside the range of 10-90%. For values see Supplementary Table S28. Error bars indicate 95% confidence intervals.

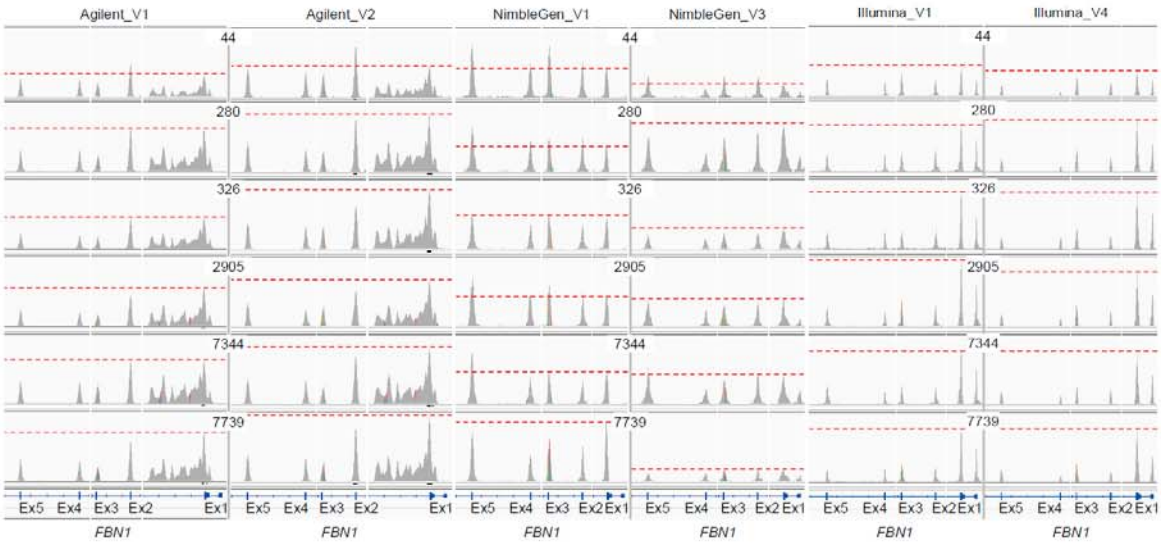



**Figure S22.** Coverage tracks displaying read depth for exons 1-5 of the *FBN1* gene. Note reduced read depth for deleted exon 1 in sample 44. Red dashed line indicates level of highest read depth of exon 1. Intronic regions between exons 1 and 2 as well as exons 3 and 4 are not shown. Coverage tracks are visualized by the Integrative Genomics Viewer (IGV) and display ranges are set to 0-500 reads for Agilent, 0-200 for NimbleGen, and 0-250 for Illumina.

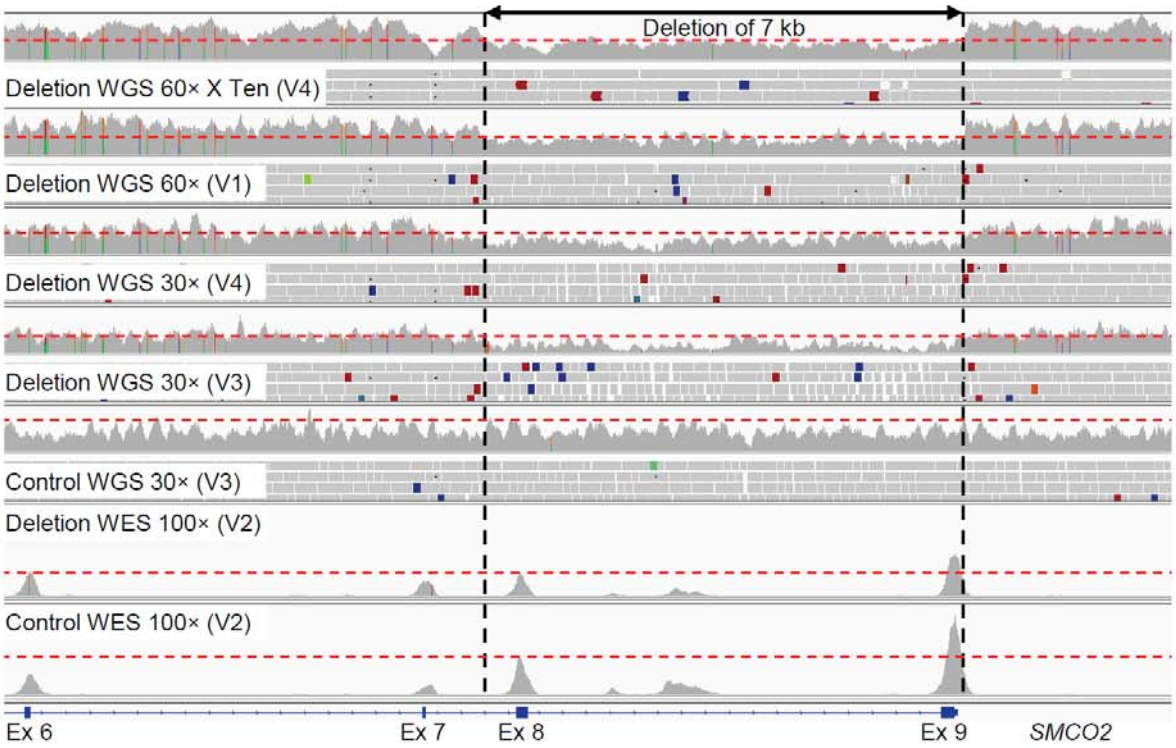



**Figure S23.** Deletion identification using WGS and WES. The genomic region of a deletion (sample 7739) detected by array CGH (4.2M NimbleGen aCGH, data not shown) is displayed by the Integrative Genomics Viewer (IGV) for both WGS and WES data. WGS was performed by V4 (HiSeq X Ten) and V1 at 60× as well as by V3 and V4 at 30×, whereas WES was carried out by V2 using Agilent SureSelect v5+UTR at 100×. For WES data, coverage track is shown only. Red dashed lines indicate level of highest read depth of exon 8. Display range of coverage tracks are set to 0-100 (V4 HiSeq X Ten), 0-60 (V1), 0-50 (V4), and 0-40 (V3) reads for WGS and 0-350 for WES.

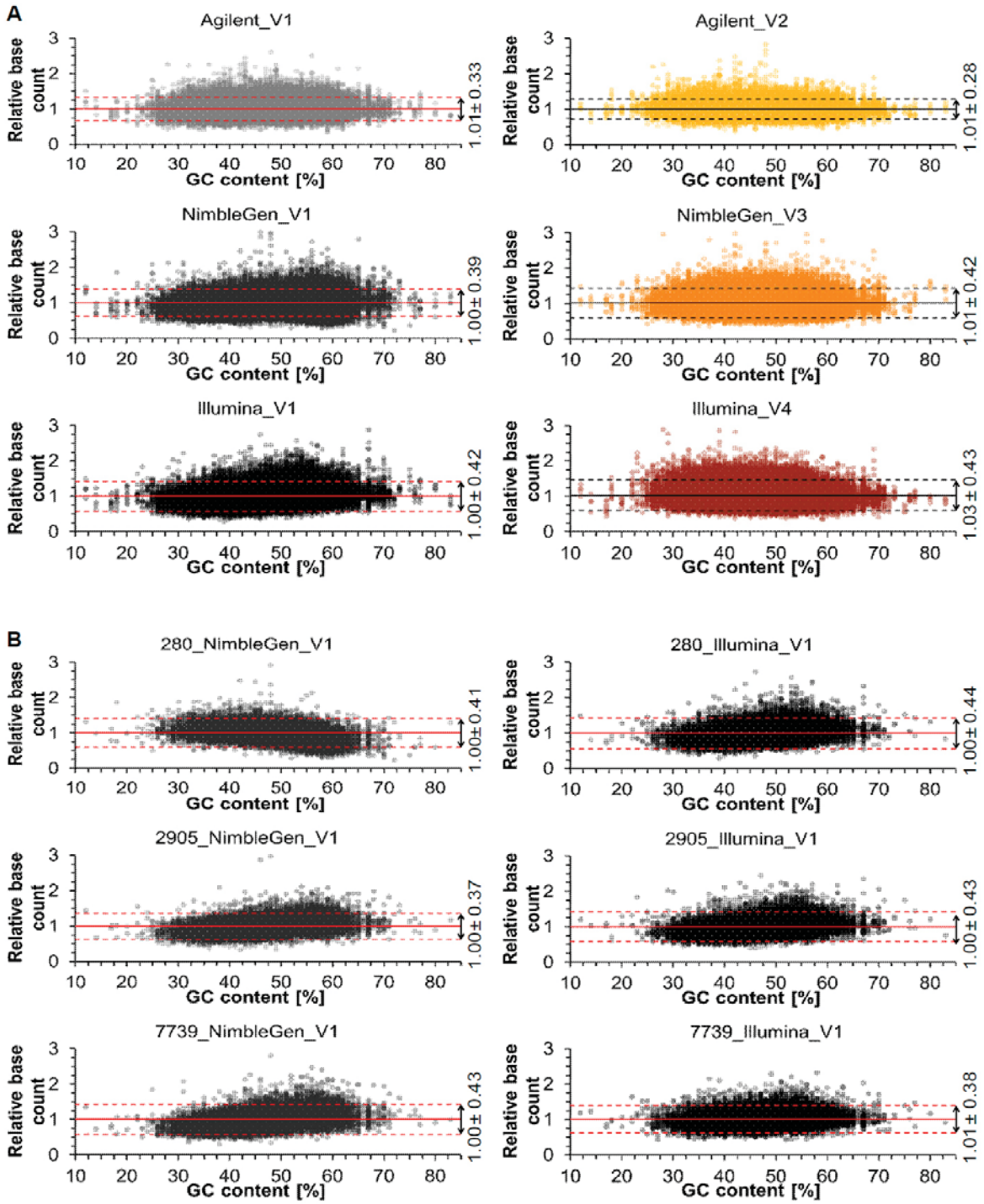



**Figure S24.** Relative WES base counts of 21'769 RefSeq exons completely (100%) covered at 20× in all 36 platform-vendor-sample combinations (plotted relative to the corresponding base counts of sample 326 and against the GC content of the respective exon). (**A**) Superposition of all five samples (44, 280, 2905, 7344, and 7739) per platform-vendor combination. (**B**) Examples for individual samples (280, 2905, and 7739) of NimbleGen and Illumina both performed by vendor V1 (V1). Note that the distribution of relative base counts for 280_NimbleGen_V1 and 7739_Illumina_V1 differ from the patterns observed in the other two exemplified samples for the same platform-vendor combination. Solid line indicates mean and dashed lines indicate an interval within which 95% of the relative base counts lie (calculated according to the Student's *t* distribution as the mean of *n* values ± critical *t* value ($t_{crit,n-1}$) × *SD* using *n* = 108'845 and 21'769 in (A) and (B), respectively, $t_{crit}$ = 1.960). Values per sample and the number of dots with relative base counts >3 not shown in this figure are given in Supplementary Table S29.

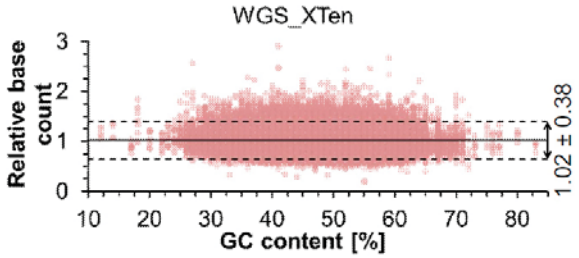



**Figure S25.** Superposition of the relative WGS base counts of 21'769 RefSeq exons (Supplementary Figure S24) of sample 7739 and four additional DNA samples plotted against the GC content of the respective exon. WGS was performed by V4 at 60× on a HiSeq X Ten system. Solid line indicates mean and dashed lines indicate an interval within which 95% of the relative base counts lie (calculated according to the Student's *t* distribution as the mean of *n* values ± critical *t* value ($t_{crit,n-1}$) × *SD* using *n* = 108'845, $t_{crit}$ = 1.960). Nine dots have relative base counts >3 and are therefore not shown in this figure.

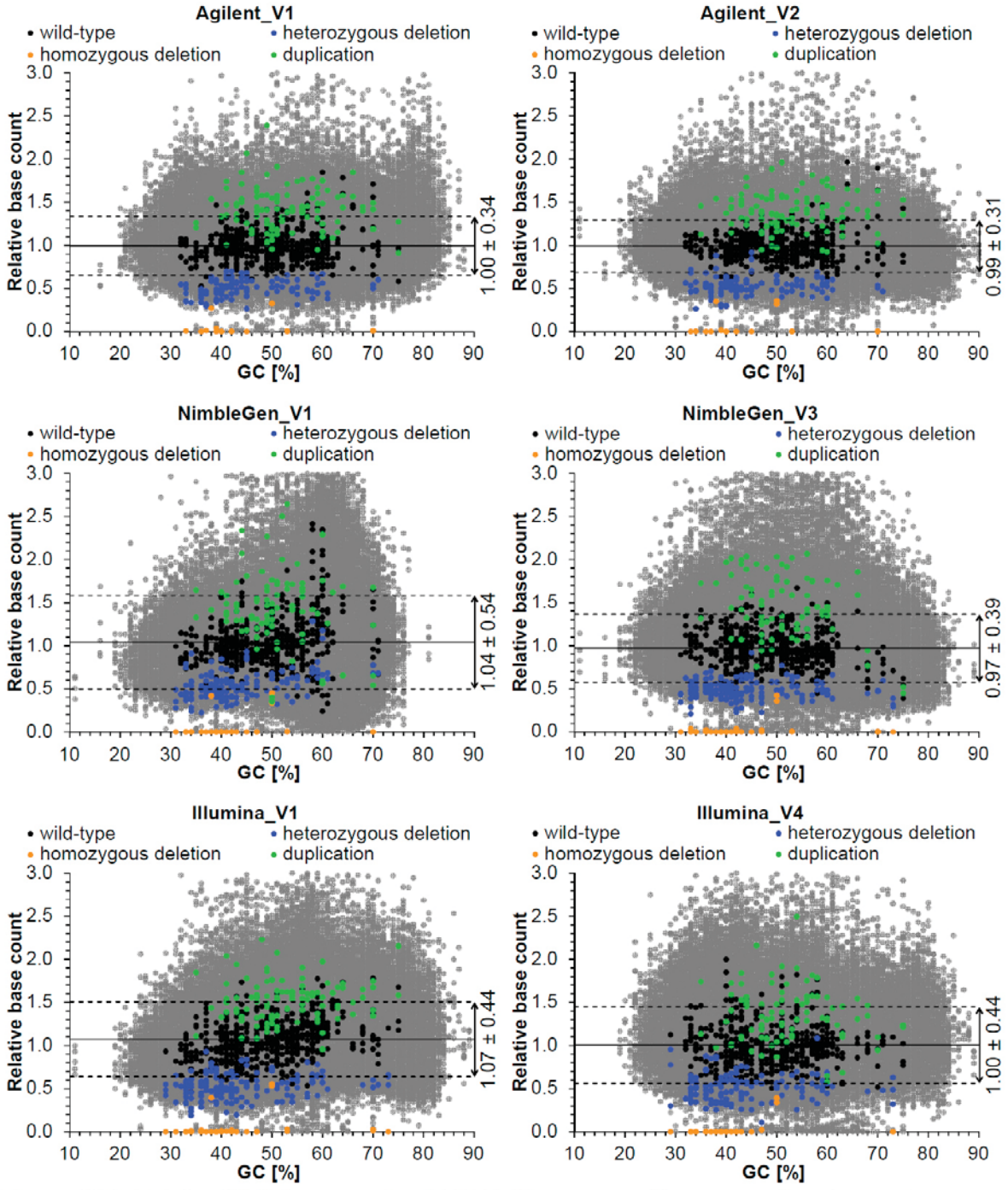



**Figure S26.** Relative WES base counts of 182 RefSeq exons with copy numbers known from array CGH (black/coloured dots according to the genotype given in Supplementary Table S21). For all six platform-vendor combinations, base counts are normalised by means of 21'769 RefSeq exons (Supplementary Figure S24) and plotted relative to the corresponding base counts of sample 326 and against the GC content of the respective exon. Relative WES base counts of ~160'000 RefSeq exons not used for normalisation with at least one base covered at 20× and a total base count of ≥1'000 in sample 326 are included and indicated (grey dots). Solid line indicates mean of the relative base counts of exons with two copies (wild-type, black dots) and dashed lines indicate an interval within which 95% of the relative base counts of these exons lie (calculated according to the Student's *t* distribution as the mean of *n* values ± critical *t* value ($t_{crit,n-1}$) × *SD* using *n* = 541, 549, 533, 534, 566, and 520 for Agilent_V1, Agilent_V2, NimbleGen_V1, NimbleGen_V3, Illumina_V1, and Illumina_V4, respectively and $t_{crit}$ = 1.960). Values per sample and the number of dots with relative base counts >3 not shown in this figure are given in Supplementary Table S30.

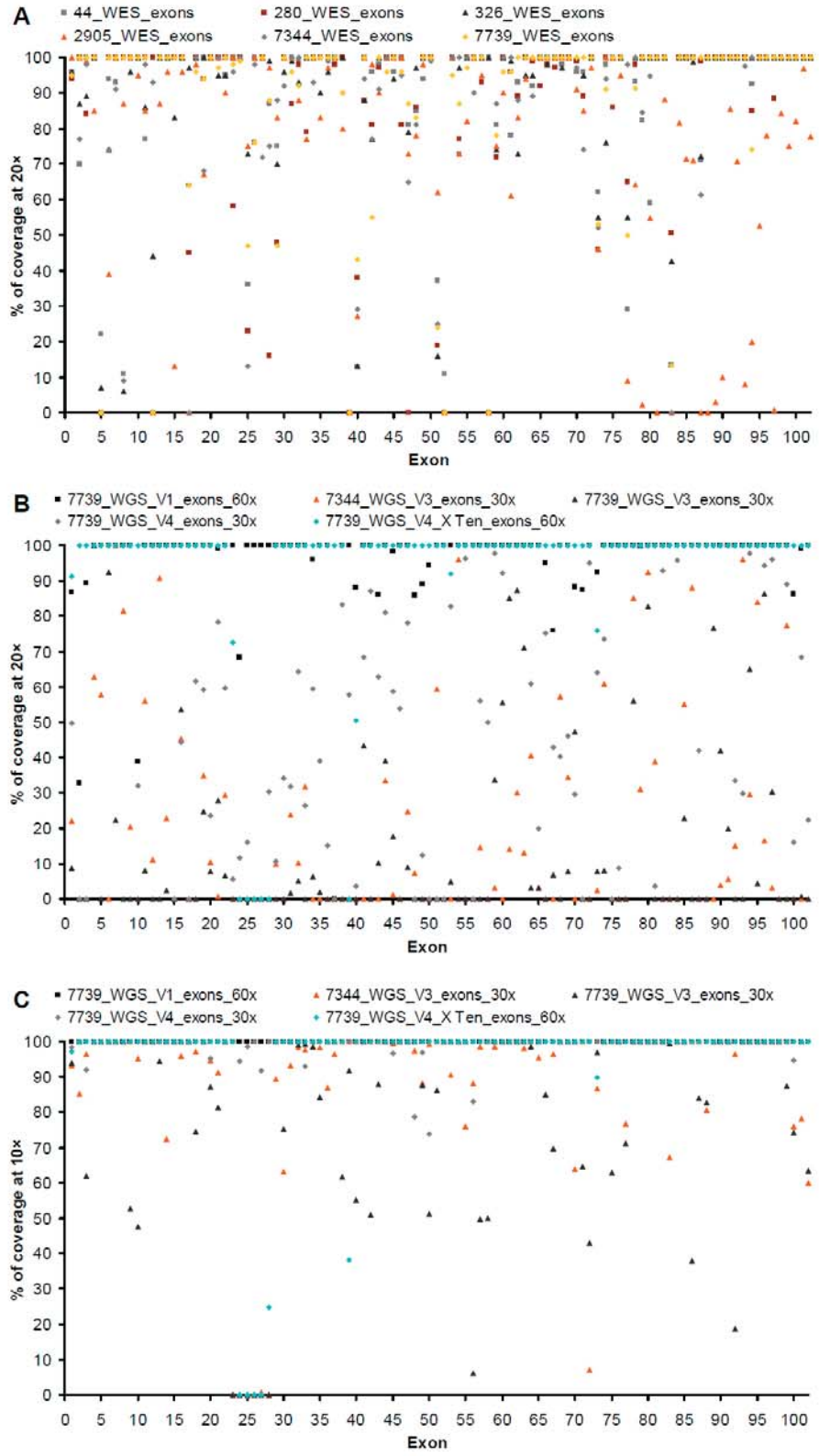



**Figure S27.** Coverage of selected RefSeq exons of 42 genes affected in monogenic disorders. (**A**) Percentage of coverage at 20× for WES data of the six DNA samples sequenced at 100× by V2 using Agilent. (**B**, **C**) Percentage of coverage at 20× (**B**) and at 10× (**C**) for WGS data of two DNA samples performed by V1 and V4 (HiSeq X Ten) at 60× as well as by V3 and V4 at 30×. Displayed RefSeq exons were covered <100% in at least one WES data set (102 out of 1'113 exons).

**Table S1.** Panel of eight selected genes and number of variants detected by Sanger sequencing in these genes in at least one of the six DNA samples used in this study. The numbers of heterozygous SNVs and indels (SNVs/indels) are summarized for each gene (all) as well as given for our region of interest (ROI, coding exons with -50-bp and +20-bp flanking intronic sequences) and UTR separately.

| | | Number of SNVs/indels | |
|---|---|---|---|
| **Genes (NM number)** | **All** | **ROI** | **UTR** |
| *COL3A1* (NM_000090) | 28/1 | 11/0 | 0/0 |
| *FBN1* (NM_000138) | 35/8 | 13/4 | 2/0 |
| *FLCN* (NM_144997) | 0/0* | 0/0* | 0/0 |
| *SLC2A10* (NM_030777) | 1/0 | 1/0 | 0/0 |
| *SMAD3* (NM_005902) | 5/0 | 1/0 | 1/0 |
| *TGFB2* (NM_001135599) | 1/0 | 0/0 | 1/0 |
| *TGFBR1* (NM_004612) | 5/2 | 1/1 | 1/0 |
| *TGFBR2* (NM_003242) | 3/0 | 3/0 | 0/0 |
| **Total** | **78/11** | **30/5** | **5/0** |

*Only two homozygous SNVs, which were excluded from analysis.

**Table S2.** Details on the design of the three exome enrichment platforms used in this study.

| Platform | **Agilent SureSelect Human All Exon kit v5+UTR** | | | **NimbleGen SeqCap EZ Exome v3+UTR** | | **Illumina Nextera Rapid Capture Expanded Exome** | |
|---|---|---|---|---|---|---|---|
| Bait length (hybridisation temperature) | 90/120 bp RNA (65°C) | | | 55-105 bp DNA (47°C) | | 95 bp DNA (58°C) | |
| Number of baits | ~881'870 | | | 2'100'000 | | 340'427 | |
| Bait density | Adjacent, partially overlapping baits across designed target region | | | High-density overlapping baits; every base covered multiple times | | Gaps between the baits | |
| Region | Designed target region | Hybridisation probes** | Padded*** | Designed target region | Hybridisation probes | Designed target region | Hybridisation probes |
| Total region (coding + UTR) | 75 Mb (50 + 25 Mb) | NA | NA | 96 Mb (64 + 32 Mb) | NA | 62 Mb (42 + 20 Mb) | NA |
| ***Calculated total region*** | ***75 Mb*** | ***73 Mb*** | ***117 Mb*** | ***96 Mb*** | ***99 Mb*** | ***62 Mb*** | ***32 Mb*** |
| Number of genes | 21'522 | NA | NA | >20'000 | NA | 20'794 | NA |
| Number of exons | 359'555 | NA | NA | NA | NA | 201'121 | NA |
| Coverage of RefSeq coding exons | 93.7%* | NA | NA | 98.4% | NA | 95.3% | NA |
| ***Calculated coverage of RefSeq coding exons*** | ***90.1%*** | ***89.6%*** | ***98.0%*** | ***98.3%*** | ***96.5%*** | ***94.9%*** | ***67.0%*** |
| ***Calculated proportion of RefSeq coding exons 100% covered*** | ***54.2%*** | ***52.8%*** | ***97.5%*** | ***98.3%*** | ***70.1%*** | ***94.0%*** | ***24.0%*** |
| ***Calculated proportion of RefSeq coding exons 0% covered*** | ***2.4%*** | ***2.6%*** | ***1.8%*** | ***1.6%*** | ***1.7%*** | ***4.9%*** | ***7.2%*** |
| ***Calculated coverage of 50-bp intronic sequences upstream of RefSeq coding exons*** | ***64.2%*** | ***63.7%*** | ***97.1%*** | ***77.5%*** | ***86.6%*** | ***8.7%*** | ***12.3%*** |
| ***Calculated coverage of 20-bp intronic sequences downstream of RefSeq coding exons*** | ***70.3%*** | ***69.7%*** | ***97.6%*** | ***93.2%*** | ***95.2%*** | ***10.6%*** | ***18.4%*** |
| Coverage of RefSeq 5'UTR and 3'UTR | 71.3%* | NA | NA | NA | NA | >88% | NA |
| Coverage of Gencode/Encode coding exons | 97.9%* | NA | NA | 96.7% | NA | 91.6% | NA |
| Coverage of Ensembl coding exons | 94.1%* | NA | NA | 99.0% | NA | 90.6% | NA |
| Coverage of CCDS exons | 98.8%* | NA | NA | 99.8% | NA | 96.0% | NA |
| Coverage of miRBase | 96.0%* | NA | NA | 98.7% | NA | >77.0% | NA |
| Input gDNA | 200 ng (low input), 3 µg (standard input) | | | 1 µg | | 50 ng | |

Information based on documents provided by companies or on correspondence with a company representative as marked by asterisk (*), except for those in bold italic which were calculated from design files downloaded from companies' websites using the SeqMonk program (http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/) and a recent release of the RefSeq database (version December 2013). Note that according to Agilent's definition the designed target region is completely covered by probes, whereas NimbleGen and Illumina define it as the region intended/designed to be enriched, which may explain lower coverage values in Agilent's designed target region (cf. also Supplementary Figure S4). **Some of Aligent's designed target regions are missing in the hybridisation probe file; ***Agilent provides an additional file with a so called padded region including designed target region as well as 100-bp flanking sequences for which sufficient coverage can be expected according to Agilent; CCDS, Consensus Coding Sequences; NA, not available; files for designed target region: Agilent: S04380219_Regions.bed, NimbleGen: track "Target Regions" in 120430_HG19_ExomeV3_UTR_EZ_HX1.bed, Illumina: nexterarapidcapture_expandedexome_targetedregions.bed; files for genomic positions of hybridisation probes: Agilent: S04380219_Probes.txt, NimbleGen: track "Tiled Regions" in 120430_HG19_ExomeV3_UTR_EZ_HX1.bed, Illumina: nexterarapidcapture_expandedexome_probes.txt); file for Agilent's padded region: S04380219_Padded.bed.

**Table S3.** Data analysis workflow and software used in this study.

| | Vendor 1 (V1)* (Agilent, NimbleGen, Illumina) | Vendor 2 (V2)* (Agilent) | Vendor 3 (V3)* (NimbleGen) | Vendor 4 (V4)* (Illumina) | In-house generated gVCF files |
|---|---|---|---|---|---|
| Pre-processing | Demultiplexing with CASAVA v1.8.2; adapter clipping; quality trimming (removal of reads containing more than one N; removal of bases or complete reads with sequencing errors; trimming of reads at 3'-end to get a minimum mean Phred quality score of 10 over a window of ten bases; reads with final length <20 bases were discarded) | Read files (FASTQ) generated from the sequencing platform via the manufacturer's proprietary software | CASAVA v1.8.2 | RTA v.1.1.2.4; CASAVA v1.8.2 | --- |
| Alignment | BWA v.0.7.5a; Picard v1.92 MarkDuplicates; GATK v.1.6-11, GATK resource bundle v2.5 for the human genome release b37: realignment around known and identified indels and recalibration of qualities using known variants to correct for biases | BWA v.0.6.2 (hg19/37); Picard v.1.98 MarkDuplicates; GATK v.1.6: local realignment of the mapped reads around potential indel sites and base quality (Phred scale) scores recalibration (GATK's covariance recalibration); SAMtools v.0.1.18: additional BAM file manipulations | ELANDv2e | BWA, v.0.5.9; SAMtools: extract mappable reads, extract on-target reads**** | BWA, v.0.7.10-r789; Picard v1.80 MarkDuplicates; GATK v3.1.1: indel realignment and base quality scores recalibration; SAMtools v0.1.19: BAM file manipulations |
| Provided/generated BAM files | Unfiltered BAM files with marked duplicates** | Unfiltered BAM files with marked duplicates** | Mapped unique reads without duplicates | Mapped unique on-target reads**** without duplicates | Unfiltered BAM files with marked duplicates** |
| Variant calling | GATK v.1.6-11, GATK resource bundle v2.5 for the human genome release b37: variant discovery and genotyping using the GATK Unified Genotyper and variant quality score recalibration using known true and false positive variant calls from the 1000 GP | GATK v.1.6: SNP and indel variants called using the GATK Unified Genotyper, dbSNP release 135 to improve quality of calls; SNP novelty determined against dbSNP release 132 | CASAVA v1.8.2; annotation using dbSNP, 1000 GP percentage, ESP percentage, and HGMD | SAMtools: extract SNVs and indels, filter SNVs and indels, and annotiation; variant databases: dbSNP and 1000 GP | GATK v. 3.1.1: Haplotype Caller |
| Region for variant calling | Platforms' designed target region | Platforms' designed target region with 100-bp flanking sequences | No restriction | Platforms' designed target region with 100-bp flanking sequences | Platforms' designed target region with 50-bp flanking sequences |
| Variants/positions in unfiltered VCF files | Quality >30, read depth >1 | Quality >30, read depth >1 | Quality ≥20, read depth >2 | Quality >3, read depth >1 | Quality >30, read depth ≥20 |
| Variant filter and recalibration settings in filtered VCF files | SNVs: ABHet >1.0, DP <10, QD <5.0, QUAL <30, and FS >60<br>Indels: DP <10, HRun >5, ReadPosRankSum <-20.0, and FS >200.0<br>Recalibration: TruthSensitivityTranches 99.00 to 99.90, 99.90 to 100+ as well as 99.90 to 100*** | SNVs: MQRankSum <-12.5, ReadPosRankSum <-8.0, QD <5.0, MQ <40.0, FS >60.0, HaplotypeScore >13<br>Indel: QD <2.0, FS >200.0, ReadPosRankSum <-20.0***<br>No recalibration | No filtered and recalibrated VCF file available | No filtered and recalibrated VCF file available | No filtered and recalibrated VCF file available |

*Information on bioinformatics pipelines of vendors V1-V4 are according to supplied documentation and information stored in provided filtered VCF files. Note that for the different analysis steps, quality and quantity of information on both data analysis pipeline and platform metrics differed among vendors. Most information on data analysis workflow was provided by V1 and V2, whereas information from V3 was only obtained upon request and the report of V4 was rather poor.

**Marked duplicates were removed from the provided/generated BAM files using Picard v1.108/1.118 MarkDuplicates.

***Nomenclature according to VCF v4.1 format.

****V4 defined on-target reads as reads within platforms' designed target region and 100-bp flanking sequences, number of aligned reads is given as total number of mapped unique reads without restriction on on-target reads (cf. Figure 1A, Supplementary Figure S13, and Supplementary Tables S5 and S24).

BWA, Burrows-Wheeler Aligner; CASAVA, Consensus Assessment of Sequence and Variation; GATK, Genome Analysis Tool Kit; RTA, real time analysis; 1000 GP, 1000 genome project; ESP, exome sequencing project; HGMD, human genome mutation database; ---, not performed; NA, not available.

**Table S4.** Enrichment efficiency for the designed target regions of the three enrichment platforms performed by the same vendor (V1).

| Agilent_V1_designed target region | 44 | 280 | 326 | 2905 | 7344 | 7739 | Mean |
|---|---|---|---|---|---|---|---|
| Total number of designed targets | 286'754 | 286'754 | 286'754 | 286'754 | 286'754 | 286'754 | **286'754** |
| Total number of aligned reads | 110'584'430 | 119'682'978 | 87'403'593 | 96'387'713 | 133'734'129 | 118'295'260 | **111'014'684** |
| Approximate mean number of reads per designed target | 270 | 289 | 211 | 233 | 322 | 290 | **269** |
| Mean read depth | 101 | 107 | 79 | 87 | 120 | 109 | **100** |
| Mean % coverage at 1× | 99.69 | 99.67 | 99.55 | 99.77 | 99.71 | 99.65 | **99.67** |
| Mean % coverage at 20× | 95.63 | 96.36 | 93.43 | 94.28 | 96.78 | 96.00 | **95.41** |
| Approximate number of reads on target* | 77'352'358 | 82'771'475 | 60'593'256 | 66'819'600 | 92'345'161 | 83'213'677 | **77'182'588** |
| Approximate % of reads on target* | 69.95 | 69.16 | 69.33 | 69.32 | 69.05 | 70.34 | **69.52** |
| Approximate number of reads on target* ±500 bp | 94'568'991 | 102'773'417 | 75'321'697 | 82'878'078 | 114'881'869 | 101'429'148 | **95'308'866** |
| Approximate % of reads off target** | 14.48 | 14.13 | 13.82 | 14.02 | 14.10 | 14.26 | **14.15** |
| **NimbleGen_V1_designed target region** | 44 | 280 | 326 | 2905 | 7344 | 7739 | **Mean** |
| Total number of designed targets | 237'172 | 237'172 | 237'172 | 237'172 | 237'172 | 237'172 | **237'172** |
| Total number of aligned reads | 186'538'676 | 129'934'094 | 122'855'749 | 151'654'787 | 143'047'403 | 159'150'546 | **148'863'543** |
| Approximate mean number of reads per designed target | 474 | 326 | 322 | 377 | 352 | 425 | **379** |
| Mean read depth | 106 | 71 | 71 | 84 | 78 | 96 | **84** |
| Mean % coverage at 1× | 91.30 | 86.62 | 89.31 | 90.29 | 90.93 | 90.98 | **89.91** |
| Mean % coverage at 20× | 81.73 | 73.46 | 77.14 | 79.15 | 79.72 | 80.53 | **78.62** |
| Approximate number of reads on target* | 112'460'374 | 77'223'935 | 76'439'364 | 89'323'164 | 83'535'838 | 100'795'033 | **89'962'951** |
| Approximate % of reads on target* | 60.29 | 59.43 | 62.22 | 58.90 | 58.40 | 63.33 | **60.43** |
| Approximate number of reads on target* ±500 bp | 141'280'259 | 98'855'834 | 97'276'602 | 113'270'989 | 105'729'484 | 125'022'431 | **113'572'600** |
| Approximate % of reads off target** | 24.26 | 23.92 | 20.82 | 25.31 | 26.09 | 21.44 | **23.71** |
| **Illumina_V1_designed target region** | 44 | 280 | 326 | 2905 | 7344 | 7739 | **Mean** |
| Total number of designed targets | 201'071 | 201'071 | 201'071 | 201'071 | 201'071 | 201'071 | **201'071** |
| Total number of aligned reads | 88'290'788 | 95'545'396 | 79'962'090 | 109'038'759 | 101'342'654 | 72'695'510 | **91'145'866** |
| Approximate mean number of reads per designed target | 197 | 218 | 180 | 242 | 223 | 165 | **204** |
| Mean read depth | 80 | 85 | 74 | 96 | 87 | 67 | **81** |
| Mean % coverage at 1× | 99.36 | 99.32 | 99.19 | 99.57 | 99.47 | 99.18 | **99.35** |
| Mean % coverage at 20× | 89.96 | 91.34 | 88.57 | 92.72 | 91.87 | 86.99 | **90.24** |
| Approximate number of reads on target* | 39'675'247 | 43'734'230 | 36'185'233 | 48'569'940 | 44'852'503 | 33'136'433 | **41'025'598** |
| Approximate % of reads on target* | 44.94 | 45.77 | 45.25 | 44.54 | 44.26 | 45.58 | **45.01** |
| Approximate number of reads on target* ±500 bp | 52'331'403 | 58'686'892 | 46'659'126 | 64'833'841 | 60'567'552 | 43'352'624 | **54'405'240** |
| Approximate % of reads off target** | 40.73 | 38.58 | 41.65 | 40.54 | 40.23 | 40.36 | **40.31** |

*Only reads within designed target region are considered as on target; **Off-target enrichment was assessed as a fraction of total aligned reads without duplicates which mapped more than 500 bp outside the designed target regions.

**Table S5.** Enrichment efficiency for the designed target regions of the three enrichment platforms performed by different vendors (V2-V4).

| Agilent_V2_designed target region | 44 | 280 | 326 | 2905 | 7344 | 7739 | Mean |
|---|---|---|---|---|---|---|---|
| Total number of designed targets | 286'754 | 286'754 | 286'754 | 286'754 | 286'754 | 286'754 | **286'754** |
| Total number of aligned reads | 172'118'729 | 160'900'330 | 157'585'695 | 129'503'393 | 156'320'331 | 156'695'689 | **155'520'695** |
| Approximate mean number of reads per designed target | 393 | 367 | 356 | 291 | 350 | 359 | **353** |
| Mean read depth | 148 | 137 | 133 | 108 | 131 | 134 | **132** |
| Mean % coverage at 1× | 99.73 | 99.72 | 99.77 | 99.86 | 99.72 | 99.73 | **99.75** |
| Mean % coverage at 20× | 97.97 | 97.79 | 97.81 | 96.78 | 97.61 | 97.67 | **97.60** |
| Approximate number of reads on target* | 112'787'851 | 105'277'451 | 102'187'865 | 83'425'965 | 100'486'466 | 102'954'139 | **101'186'623** |
| Approximate % of reads on target* | 65.53 | 65.43 | 64.85 | 64.42 | 64.28 | 65.70 | **65.06** |
| Approximate number of reads on target* ±500 bp | 140'118'470 | 133'378'927 | 129'248'040 | 106'320'714 | 128'338'725 | 128'600'703 | **127'667'597** |
| Approximate % of reads off target** | 18.59 | 17.10 | 17.98 | 17.90 | 17.90 | 17.93 | **17.91** |
| **NimbleGen_V3_designed target region** | 44 | 280 | 326 | 2905 | 7344 | 7739 | **Mean** |
| Total number of designed targets | 237'172 | 237'172 | 237'172 | 237'172 | 237'172 | 237'172 | **237'172** |
| Total number of aligned reads*** | 72'899'287 | 158'324'581 | 61'382'269 | 90'817'153 | 121'277'235 | 41'985'030 | **91'114'259** |
| Approximate mean number of reads per designed target | 203 | 446 | 172 | 250 | 334 | 115 | **253** |
| Mean read depth | 46 | 94 | 39 | 55 | 72 | 26 | **55** |
| Mean % coverage at 1× | 98.56 | 98.70 | 98.48 | 98.74 | 98.70 | 98.36 | **98.59** |
| Mean % coverage at 20× | 92.12 | 95.51 | 89.17 | 91.48 | 94.60 | 71.83 | **89.12** |
| Approximate number of reads on target* | 48'213'734 | 105'669'383 | 40'763'348 | 59'221'672 | 79'236'343 | 27'159'248 | **60'043'955** |
| Approximate % of reads on target* | 66.14 | 66.74 | 66.41 | 65.21 | 65.33 | 64.69 | **65.90** |
| Approximate number of reads on target* ±500 bp | 64'166'634 | 139'984'744 | 54'364'051 | 79'815'713 | 106'988'793 | 37'100'565 | **80'403'417** |
| Approximate % of reads off target** | 11.98 | 11.58 | 11.43 | 12.11 | 11.78 | 11.63 | **11.76** |
| **Illumina_V4_designed target region** | 44 | 280 | 326 | 2905 | 7344 | 7739 | **Mean** |
| Total number of designed targets | 201'071 | 201'071 | 201'071 | 201'071 | 201'071 | 201'071 | **201'071** |
| Total number of aligned reads*** | 53'245'236 | 45'562'196 | 56'364'527 | 57'138'610 | 55'996'771 | 47'498'550 | **52'634'315** |
| Approximate mean number of reads per designed target | 137 | 114 | 140 | 143 | 139 | 121 | **132** |
| Mean read depth | 54 | 46 | 57 | 57 | 54 | 48 | **53** |
| Mean % coverage at 1× | 98.49 | 98.32 | 98.50 | 98.63 | 98.58 | 98.41 | **98.49** |
| Mean % coverage at 20× | 85.85 | 81.45 | 86.24 | 86.00 | 82.44 | 82.44 | **84.07** |
| Approximate number of reads on target* | 27'610'149 | 22'994'522 | 28'087'064 | 28'707'234 | 27'898'446 | 24'301'298 | **26'599'785** |
| Approximate % of reads on target* | 51.85 | 50.47 | 49.83 | 50.24 | 49.82 | 51.16 | **50.54** |
| Approximate number of reads on target* ±500 bp | 33'196'467 | 27'625'088 | 33'570'343 | 34'504'989 | 33'608'122 | 29'249'735 | **31'959'124** |
| Approximate % of reads off target** | 37.65 | 39.37 | 40.44 | 39.61 | 39.98 | 38.42 | **39.28** |

*Only reads within designed target region are considered as on target; **Off-target enrichment was assessed as a fraction of total aligned reads without duplicates which mapped more than 500 bp outside the designed target regions; ***only unique reads (cf. Supplementary Table S3).

**Table S6.** Enrichment efficiency of the three enrichment platforms performed by the same vendor (V1) for RefSeq exons.

| Agilent_V1_RefSeq_all | 44 | 280 | 326 | 2905 | 7344 | 7739 | Mean |
|---|---|---|---|---|---|---|---|
| Total number of exons | 232'619 | 232'619 | 232'619 | 232'619 | 232'619 | 232'619 | 232'619 |
| Approximate number of reads on exons | 62'700'606 | 67'193'698 | 49'197'375 | 54'225'588 | 75'061'028 | 67'190'204 | 62'594'750 |
| Approximate mean number of reads per exon | 270 | 289 | 211 | 233 | 323 | 289 | 269 |
| Mean read depth | 90 | 95 | 70 | 77 | 106 | 96 | 89 |
| Mean % coverage at 1× | 91.18 | 91.35 | 90.47 | 90.96 | 91.49 | 91.19 | 91.11 |
| Mean % coverage at 20× | 83.57 | 84.27 | 81.67 | 82.43 | 84.74 | 83.79 | 83.41 |
| **NimbleGen_V1_RefSeq_all** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of exons | 232'619 | 232'619 | 232'619 | 232'619 | 232'619 | 232'619 | 232'619 |
| Approximate number of reads on exons | 88'680'868 | 60'566'348 | 60'014'994 | 70'443'258 | 65'994'615 | 79'544'377 | 70'874'077 |
| Approximate mean number of reads per exon | 381 | 260 | 258 | 303 | 284 | 342 | 305 |
| Mean read depth | 111 | 73 | 74 | 87 | 82 | 101 | 88 |
| Mean % coverage at 1× | 88.92 | 83.36 | 86.01 | 87.67 | 88.40 | 88.15 | 87.09 |
| Mean % coverage at 20× | 77.45 | 68.82 | 72.83 | 75.00 | 75.72 | 76.31 | 74.36 |
| **Illumina_V1_RefSeq_all** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of exons | 232'619 | 232'619 | 232'619 | 232'619 | 232'619 | 232'619 | 232'619 |
| Approximate number of reads on exons | 43'672'521 | 48'168'416 | 39'641'712 | 53'453'760 | 49'393'423 | 36'401'395 | 45'121'871 |
| Approximate mean number of reads per exon | 188 | 207 | 170 | 230 | 212 | 156 | 194 |
| Mean read depth | 73 | 78 | 68 | 88 | 80 | 61 | 75 |
| Mean % coverage at 1× | 94.40 | 94.43 | 93.85 | 95.34 | 95.11 | 93.62 | 94.46 |
| Mean % coverage at 20× | 81.61 | 82.94 | 80.23 | 84.31 | 83.48 | 78.81 | 81.90 |
| **Agilent_V1_RefSeq_coding region** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of exons | 202'044 | 202'044 | 202'044 | 202'044 | 202'044 | 202'044 | 202'044 |
| Approximate number of reads on exons | 39'019'740 | 41'231'642 | 30'210'489 | 33'413'675 | 46'161'327 | 41'534'098 | 38'595'162 |
| Approximate mean number of reads per exon | 193 | 204 | 150 | 165 | 228 | 206 | 191 |
| Mean read depth | 106 | 113 | 82 | 91 | 126 | 114 | 105 |
| Mean % coverage at 1× | 98.88 | 98.89 | 98.74 | 98.93 | 98.93 | 98.86 | 98.87 |
| Mean % coverage at 15× | 96.34 | 96.70 | 95.22 | 95.73 | 96.94 | 96.41 | 96.22 |
| Mean % coverage at 20× | 94.89 | 95.52 | 92.97 | 93.72 | 95.96 | 95.13 | 94.70 |
| **NimbleGen_V1_RefSeq_coding region** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of exons | 202'044 | 202'044 | 202'044 | 202'044 | 202'044 | 202'044 | 202'044 |
| Approximate number of reads on exons | 44'148'941 | 28'580'556 | 28'882'087 | 34'833'826 | 32'716'041 | 40'243'688 | 34'900'857 |
| Approximate mean number of reads per exon | 219 | 141 | 143 | 172 | 162 | 199 | 173 |
| Mean read depth | 118 | 77 | 78 | 92 | 87 | 108 | 93 |
| Mean % coverage at 1× | 91.66 | 86.20 | 89.39 | 90.39 | 91.22 | 91.35 | 90.04 |
| Mean % coverage at 15× | 83.76 | 74.96 | 79.37 | 81.45 | 82.41 | 82.77 | 80.79 |
| Mean % coverage at 20× | 82.23 | 73.03 | 77.42 | 79.64 | 80.55 | 81.11 | 79.00 |
| **Illumina_V1_RefSeq_coding region** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of exons | 202'044 | 202'044 | 202'044 | 202'044 | 202'044 | 202'044 | 202'044 |
| Approximate number of reads on exons | 25'832'596 | 28'353'998 | 23'513'955 | 31'645'766 | 29'003'715 | 21'497'098 | 26'641'188 |
| Approximate mean number of reads per exon | 128 | 140 | 116 | 157 | 144 | 106 | 132 |
| Mean read depth | 77 | 83 | 72 | 93 | 84 | 64 | 79 |
| Mean % coverage at 1× | 97.64 | 97.64 | 97.38 | 98.10 | 97.94 | 97.30 | 97.67 |
| Mean % coverage at 15× | 90.05 | 91.13 | 88.85 | 92.05 | 91.57 | 88.05 | 90.28 |
| Mean % coverage at 20× | 86.62 | 88.18 | 85.31 | 89.47 | 88.61 | 83.75 | 86.99 |
| **Agilent_V1_RefSeq_5'UTR** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of exons | 54'367 | 54'367 | 54'367 | 54'367 | 54'367 | 54'367 | 54'367 |
| Approximate number of reads on exons | 6'377'126 | 6'675'046 | 4'884'380 | 5'431'411 | 7'389'371 | 6'520'103 | 6'212'906 |
| Approximate mean number of reads per exon | 117 | 123 | 90 | 100 | 136 | 120 | 114 |
| Mean read depth | 75 | 79 | 57 | 64 | 88 | 78 | 74 |
| Mean % coverage at 1× | 76.09 | 76.42 | 73.59 | 75.17 | 76.82 | 75.95 | 75.67 |
| Mean % coverage at 20× | 59.76 | 60.43 | 57.51 | 58.69 | 60.80 | 59.30 | 59.41 |
| **NimbleGen_V1_RefSeq_5'UTR** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of exons | 54'367 | 54'367 | 54'367 | 54'367 | 54'367 | 54'367 | 54'367 |
| Approximate number of reads on exons | 8'353'576 | 5'446'837 | 5'498'776 | 6'667'355 | 6'228'849 | 7'488'157 | 6'613'925 |
| Approximate mean number of reads per exon | 154 | 100 | 101 | 123 | 115 | 138 | 122 |
| Mean read depth | 77 | 50 | 51 | 61 | 58 | 70 | 61 |
| Mean % coverage at 1× | 72.68 | 65.89 | 68.41 | 71.05 | 71.97 | 71.52 | 70.25 |
| Mean % coverage at 20× | 56.37 | 48.31 | 51.72 | 54.04 | 54.53 | 55.22 | 53.37 |
| **Illumina_V1_RefSeq_5'UTR** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of exons | 54'367 | 54'367 | 54'367 | 54'367 | 54'367 | 54'367 | 54'367 |
| Approximate number of reads on exons | 5'909'735 | 6'236'696 | 5'328'908 | 7'026'514 | 6'536'708 | 4'863'018 | 5'983'597 |
| Approximate mean number of reads per exon | 109 | 115 | 98 | 129 | 120 | 89 | 110 |
| Mean read depth | 58 | 60 | 54 | 68 | 63 | 48 | 59 |
| Mean % coverage at 1× | 86.76 | 86.66 | 85.53 | 88.56 | 88.26 | 85.03 | 86.80 |
| Mean % coverage at 20× | 64.53 | 65.10 | 62.36 | 67.45 | 66.56 | 61.05 | 64.51 |
| **Agilent_V1_RefSeq_3'UTR** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of exons | 38'600 | 38'600 | 38'600 | 38'600 | 38'600 | 38'600 | 38'600 |
| Approximate number of reads on exons | 22'781'870 | 25'188'734 | 18'416'436 | 20'179'031 | 28'116'725 | 25'049'443 | 23'288'706 |
| Approximate mean number of reads per exon | 590 | 653 | 477 | 523 | 728 | 649 | 603 |
| Mean read depth | 76 | 82 | 60 | 66 | 92 | 82 | 76 |
| Mean % coverage at 1× | 82.58 | 83.15 | 81.69 | 82.32 | 83.36 | 82.82 | 82.65 |
| Mean % coverage at 20× | 70.26 | 71.20 | 68.70 | 69.37 | 71.76 | 70.56 | 70.31 |
| **NimbleGen_V1_RefSeq_3'UTR** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of exons | 38'600 | 38'600 | 38'600 | 38'600 | 38'600 | 38'600 | 38'600 |
| Approximate number of reads on exons | 43'473'480 | 31'653'358 | 30'648'749 | 34'766'931 | 32'483'625 | 38'331'765 | 35'226'318 |
| Approximate mean number of reads per exon | 1126 | 820 | 794 | 901 | 842 | 993 | 913 |
| Mean read depth | 116 | 80 | 79 | 91 | 86 | 103 | 92 |
| Mean % coverage at 1× | 90.39 | 85.73 | 86.86 | 89.62 | 89.80 | 88.68 | 88.51 |
| Mean % coverage at 20× | 75.98 | 69.25 | 72.10 | 74.00 | 74.00 | 74.77 | 73.35 |
| **Illumina_V1_RefSeq_3'UTR** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of exons | 38'600 | 38'600 | 38'600 | 38'600 | 38'600 | 38'600 | 38'600 |
| Approximate number of reads on exons | 14'879'463 | 16'813'223 | 13'545'316 | 18'384'906 | 17'152'881 | 12'513'243 | 15'548'172 |
| Approximate mean number of reads per exon | 385 | 436 | 351 | 476 | 444 | 324 | 403 |
| Mean read depth | 51 | 57 | 47 | 63 | 58 | 43 | 53 |
| Mean % coverage at 1× | 88.02 | 88.35 | 86.51 | 90.25 | 89.79 | 86.13 | 88.17 |
| Mean % coverage at 20× | 63.36 | 66.11 | 60.35 | 67.72 | 66.96 | 59.38 | 63.98 |

**Table S7.** Enrichment efficiency of the three enrichment platforms performed by the same vendor (V1) for intronic sequences (regions) in our region of interest (RefSeq coding exons and -50-bp and +20-bp flanking intronic sequences).

| **Agilent_V1_RefSeq_coding_-50 bp** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
|---|---|---|---|---|---|---|---|
| Total number of regions | 199'107 | 199'107 | 199'107 | 199'107 | 199'107 | 199'107 | **199'107** |
| Approximate number of reads on regions | 7'920'308 | 8'502'572 | 6'214'050 | 6'888'265 | 9'510'776 | 8'449'249 | **7'914'203** |
| Approximate mean number of reads per region | 40 | 43 | 31 | 35 | 48 | 42 | **40** |
| Mean read depth | 80 | 85 | 62 | 69 | 96 | 85 | **79** |
| Mean % coverage at 15× | 91.30 | 92.98 | 89.04 | 90.12 | 93.85 | 91.69 | **91.50** |
| Mean % coverage at 20× | 87.17 | 89.56 | 83.61 | 85.22 | 90.96 | 87.95 | **87.41** |
| **NimbleGen_V1_RefSeq_coding_-50 bp** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of regions | 199'107 | 199'107 | 199'107 | 199'107 | 199'107 | 199'107 | **199'107** |
| Approximate number of reads on regions | 9'159'017 | 6'170'791 | 6'173'059 | 7'270'276 | 6'809'552 | 8'260'991 | **7'307'281** |
| Approximate mean number of reads per region | 46 | 31 | 31 | 37 | 34 | 41 | **37** |
| Mean read depth | 92 | 62 | 62 | 73 | 68 | 83 | **73** |
| Mean % coverage at 15× | 83.22 | 75.16 | 78.93 | 80.70 | 81.39 | 82.00 | **80.23** |
| Mean % coverage at 20× | 81.15 | 72.49 | 76.15 | 78.06 | 78.56 | 79.63 | **77.67** |
| **Illumina_V1_RefSeq_coding_-50 bp** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of regions | 199'107 | 199'107 | 199'107 | 199'107 | 199'107 | 199'107 | **199'107** |
| Approximate number of reads on regions | 3'853'422 | 4'356'279 | 3'427'392 | 4'798'041 | 4'455'867 | 3'175'805 | **4'011'134** |
| Approximate mean number of reads per region | 19 | 22 | 17 | 24 | 22 | 16 | **20** |
| Mean read depth | 39 | 44 | 34 | 48 | 45 | 32 | **40** |
| Mean % coverage at 15× | 69.47 | 74.60 | 64.63 | 76.83 | 75.49 | 63.12 | **70.69** |
| Mean % coverage at 20× | 59.39 | 65.20 | 54.87 | 68.03 | 66.04 | 52.51 | **61.01** |
| **Agilent_V1_RefSeq_coding_+20 bp** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of regions | 197'564 | 197'564 | 197'564 | 197'564 | 197'564 | 197'564 | **197'564** |
| Approximate number of reads on regions | 3'518'506 | 3'754'173 | 2'751'585 | 3'037'132 | 4'200'888 | 3'771'193 | **3'505'579** |
| Approximate mean number of reads per region | 18 | 19 | 14 | 15 | 21 | 19 | **18** |
| Mean read depth | 89 | 95 | 70 | 77 | 106 | 95 | **89** |
| Mean % coverage at 15× | 93.40 | 94.42 | 91.40 | 92.21 | 95.12 | 93.56 | **93.35** |
| Mean % coverage at 20× | 90.15 | 91.73 | 86.87 | 88.15 | 92.88 | 90.62 | **90.07** |
| **NimbleGen_V1_RefSeq_coding_+20 bp** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of regions | 197'564 | 197'564 | 197'564 | 197'564 | 197'564 | 197'564 | **197'564** |
| Approximate number of reads on regions | 4'278'498 | 2'817'235 | 2'843'002 | 3'359'239 | 3'161'171 | 3'852'469 | **3'385'269** |
| Approximate mean number of reads per region | 22 | 14 | 14 | 17 | 16 | 19 | **17** |
| Mean read depth | 108 | 71 | 72 | 85 | 80 | 97 | **86** |
| Mean % coverage at 15× | 83.23 | 74.56 | 78.73 | 80.88 | 81.70 | 82.13 | **80.21** |
| Mean % coverage at 20× | 81.51 | 72.47 | 76.56 | 78.81 | 79.55 | 80.24 | **78.19** |
| **Illumina_V1_RefSeq_coding_+20 bp** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of regions | 197'564 | 197'564 | 197'564 | 197'564 | 197'564 | 197'564 | **197'564** |
| Approximate number of reads on regions | 1'766'131 | 1'992'962 | 1'992'962 | 2'203'580 | 2'039'491 | 1'459'744 | **1'909'145** |
| Approximate mean number of reads per region | 9 | 10 | 10 | 11 | 10 | 7 | **10** |
| Mean read depth | 45 | 50 | 50 | 56 | 52 | 37 | **48** |
| Mean % coverage at 15× | 76.66 | 80.59 | 73.27 | 82.81 | 81.58 | 71.35 | **77.71** |
| Mean % coverage at 20× | 67.63 | 72.55 | 64.19 | 75.46 | 73.50 | 61.19 | **69.09** |

**Table S8.** Uniformity of the coverage of RefSeq coding exons.

| **Agilent_V1_RefSeq_coding region_uniformity** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
|---|---|---|---|---|---|---|---|
| Total number of exons | 202'044 | 202'044 | 202'044 | 202'044 | 202'044 | 202'044 | **202'044** |
| Mean read depth | 106 | 113 | 82 | 91 | 126 | 114 | **105** |
| Exons with 100% coverage at 20× | 175'281 | 179'977 | 166'032 | 169'766 | 183'167 | 176'838 | **175'177** |
| Exons with 100% coverage at 20× [%] | 86.75 | 89.08 | 82.18 | 84.02 | 90.66 | 87.52 | **86.70** |
| Exons with average read depth within ±70% of mean read depth | 160'167 | 161'721 | 161'019 | 159'423 | 160'356 | 160'068 | **160'459** |
| Exons with average read depth within ±70% of mean read depth [%] | 79.27 | 80.04 | 79.70 | 78.91 | 79.37 | 79.22 | **79.42** |
| **NimbleGen_V1_RefSeq_coding region_uniformity** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of exons | 202'044 | 202'044 | 202'044 | 202'044 | 202'044 | 202'044 | **202'044** |
| Mean read depth | 118 | 77 | 78 | 92 | 87 | 108 | **93** |
| Exons with 100% coverage at 20× | 153'113 | 134'901 | 142'875 | 147'416 | 149'468 | 149'806 | **146'263** |
| Exons with 100% coverage at 20× [%] | 75.78 | 66.77 | 70.71 | 72.96 | 73.98 | 74.15 | **72.39** |
| Exons with average read depth within ±70% of mean read depth | 131'419 | 111'898 | 125'842 | 126'789 | 131'882 | 129'236 | **126'178** |
| Exons with average read depth within ±70% of mean read depth [%] | 65.04 | 55.38 | 62.28 | 62.75 | 65.27 | 63.96 | **62.45** |
| **Illumina_V1_RefSeq_coding region_uniformity** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of exons | 202'044 | 202'044 | 202'044 | 202'044 | 202'044 | 202'044 | **202'044** |
| Mean read depth | 77 | 83 | 72 | 93 | 84 | 64 | **79** |
| Exons with 100% coverage at 20× | 124'912 | 134'570 | 117'424 | 141'082 | 136'283 | 111'031 | **127'550** |
| Exons with 100% coverage at 20× [%] | 61.82 | 66.60 | 58.12 | 69.83 | 67.45 | 54.95 | **63.13** |
| Exons with average read depth within ±70% of mean read depth | 150'075 | 149'421 | 154'065 | 149'234 | 149'318 | 151'963 | **150'679** |
| Exons with average read depth within ±70% of mean read depth [%] | 74.28 | 73.95 | 76.25 | 73.86 | 73.90 | 75.21 | **74.58** |
| **Agilent_V2_RefSeq_coding region_uniformity** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of exons | 202'044 | 202'044 | 202'044 | 202'044 | 202'044 | 202'044 | **202'044** |
| Mean read depth | 156 | 144 | 140 | 114 | 137 | 142 | **139** |
| Exons with 100% coverage at 20× | 188'839 | 188'779 | 188'675 | 183'648 | 188'298 | 187'339 | **187'596** |
| Exons with 100% coverage at 20× [%] | 93.46 | 93.43 | 93.38 | 90.90 | 93.20 | 92.72 | **92.85** |
| Exons with average read depth within ±70% of mean read depth | 161'636 | 162'627 | 162'929 | 161'077 | 162'061 | 161'861 | **162'032** |
| Exons with average read depth within ±70% of mean read depth [%] | 80.00 | 80.49 | 80.64 | 79.72 | 80.21 | 80.11 | **80.20** |
| **NimbleGen_V3_RefSeq_coding region_uniformity** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of exons | 202'044 | 202'044 | 202'044 | 202'044 | 202'044 | 202'044 | **202'044** |
| Mean read depth | 51 | 104 | 42 | 59 | 79 | 28 | **61** |
| Exons with 100% coverage at 20× | 184'605 | 193'678 | 175'023 | 181'855 | 191'318 | 114'754 | **173'539** |
| Exons with 100% coverage at 20× [%] | 91.37 | 95.86 | 86.63 | 90.01 | 94.69 | 56.80 | **85.89** |
| Exons with average read depth within ±70% of mean read depth | 192'342 | 182'511 | 192'255 | 185'321 | 187'782 | 190'225 | **188'406** |
| Exons with average read depth within ±70% of mean read depth [%] | 95.20 | 90.33 | 95.16 | 91.72 | 92.94 | 94.15 | **93.25** |
| **Illumina_V4_RefSeq_coding region_uniformity** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of exons | 202'044 | 202'044 | 202'044 | 202'044 | 202'044 | 202'044 | **202'044** |
| Mean read depth | 53 | 45 | 55 | 55 | 52 | 46 | **51** |
| Exons with 100% coverage at 20× | 112'625 | 95'971 | 112'640 | 114'902 | 113'604 | 100'317 | **108'343** |
| Exons with 100% coverage at 20× [%] | 55.74 | 47.50 | 55.75 | 56.87 | 56.23 | 49.65 | **53.62** |
| Exons with average read depth within ±70% of mean read depth | 152'575 | 154'422 | 153'579 | 152'341 | 152'382 | 152'967 | **153'044** |
| Exons with average read depth within ±70% of mean read depth [%] | 75.52 | 76.43 | 76.01 | 75.40 | 75.42 | 75.71 | **75.75** |

**Table S9.** Coverage of RefSeq coding exons by whole genome sequencing (WGS).

| WGS_RefSeq_coding region | 7739_V1 (60×) | 7344_V3 (30×) | 7739_V3 (30×) | 7739_V4 (30×) | 7739_V4_X Ten (60×) |
|---|---|---|---|---|---|
| Total number of exons | 202'044 | 202'044 | 202'044 | 202'044 | 202'044 |
| Approximate number of reads on exons | 11'674'285 | 5'461'115 | 5'113'945 | 7'708'299 | 22'846'600 |
| Approximate mean number of reads per exon | 58 | 27 | 25 | 38 | 75 |
| Mean read depth | 33 | 16 | 15 | 22 | 65 |
| Mean % coverage at 10× | 99.63 | 91.60 | 88.56 | 98.37 | 99.02 |
| Mean % coverage at 20× | 97.57 | 21.49 | 14.83 | 64.21 | 98.68 |
| Exons with average read depth within ±70% of mean read depth [%] | 99.33 | 97.54 | 97.27 | 98.84 | 98.14 |
| Exons with <100% coverage at 20× [%] | 12.55 | 96.92 | 98.21 | 69.81 | 1.84 |
| Exons with <100% coverage at 15× [%] | 2.65 | 79.05 | 85.42 | 37.26 | 1.55 |
| Exons with <100% coverage at 10× [%] | 0.74 | 32.17 | 40.45 | 9.12 | 1.31 |

X Ten, HiSeq X Ten system.

**Table S10.** Proportions of RefSeq coding exons not completely (100%) covered at 20× neither by using the three platforms alone nor in any combination. In the last row, the proportions of exons completely covered by all six platform-vendor combinations (shared exons covered at 20×) are given in italics as well. If not otherwise indicated, data of all corresponding vendors are included.

| Platform(s) | 44 | 280 | 326 | 2905 | 7344 | 7739 | Mean* |
|---|---|---|---|---|---|---|---|
| Agilent (vendor V1) | 13.25% | 10.92% | 17.82% | 15.98% | 9.34% | 12.48% | **13.30±3.31%** |
| NimbleGen (vendor V1) | 24.22% | 33.23% | 29.29% | 27.04% | 26.02% | 25.85% | **27.61±3.38%** |
| Illumina (vendor V1) | 38.18% | 33.40% | 41.88% | 30.17% | 32.55% | 45.05% | **36.87±6.11%** |
| Agilent (vendor V2) | 6.54% | 6.57% | 6.62% | 9.10% | 6.80% | 7.28% | **7.15±1.05%** |
| NimbleGen (vendor V3) | 8.63% | 4.14% | 13.37% | 9.99% | 5.31% | 43.20% | **14.11±15.36%** |
| Illumina (vendor V4) | 44.26% | 52.50% | 44.25% | 43.13% | 43.77% | 50.35% | **46.38±4.19%** |
| Agilent (vendors V1 and V2) | 6.36% | 6.21% | 6.54% | 8.51% | 6.19% | 6.87% | **6.78±0.93%** |
| NimbleGen (vendors V1 and V3) | 6.33% | 3.25% | 10.24% | 7.92% | 4.19% | 20.74% | **8.78±6.7%** |
| Illumina (vendors V1 and V4) | 31.20% | 29.90% | 33.34% | 25.96% | 27.38% | 37.12% | **30.82±4.27%** |
| Agilent and NimbleGen | 2.08% | 1.78% | 2.49% | 2.41% | 1.98% | 3.19% | **2.32±0.53%** |
| Agilent and Illumina | 2.95% | 2.83% | 3.10% | 3.32% | 2.63% | 3.48% | **3.05±0.33%** |
| NimbleGen and Illumina | 3.45% | 2.13% | 4.91% | 3.47% | 2.48% | 8.37% | **4.13±2.40%** |
| All three platforms | 1.58% | 1.39% | 1.78% | 1.55% | 1.50% | 2.11% | **1.65±0.27%** |
| *Shared exons covered at 20× [%]* | *32.24%* | *25.60%* | *27.68%* | *33.01%* | *34.15%* | *19.65%* | ***28.72±5.81%*** |

*Indicated ranges for mean values (±) represent 95% confidence intervals.

**Table S11.** Coverage of RefSeq coding exons of ~7600 genes targeted by the Accuracy and Content Enhanced (ACEv2) clinical exome platform of Personalis (www.personalis.com) achieved using WES at 100×, WGS at 30× and 60×, and Personalis ACEv2 at 60×.

| | Total exons | Mean read depth | Mean coverage at 20× [%] | Exons with <100% coverage at 20× [%] | Exons with <100% coverage at 15× [%] | Exons with <100% coverage at 10× [%] |
|---|---|---|---|---|---|---|
| 7739_Agilent_V1 | 99'236 | 115 | 95.98 | 11.50 | 7.50 | 4.36 |
| 7739_Agilent_V2 | 99'236 | 143 | 97.41 | 6.31 | 4.10 | 2.55 |
| 7739_NimbleGen_V1 | 99'236 | 109 | 82.59 | 24.26 | 22.35 | 20.33 |
| 7739_NimbleGen_V3 | 99'236 | 28 | 81.37 | 41.69 | 20.51 | 8.29 |
| 7739_Illumina_V1 | 99'236 | 63 | 83.79 | 46.07 | 35.60 | 23.76 |
| 7739_Illumina_V4 | 99'236 | 47 | 81.79 | 47.52 | 34.74 | 21.62 |
| 7739_WGS_V3_30× | 99'236 | 15 | 15.47 | 97.98 | 85.04 | 39.23 |
| 7739_WGS_V4_30× | 99'236 | 22 | 65.18 | 69.35 | 36.09 | 8.55 |
| 7739_WGS_V1_60× | 99'236 | 34 | 97.72 | 11.83 | 2.39 | 0.71 |
| 7739_WGS_XTen_V4_60× (non-PCR-free) | 99'236 | 66 | 98.91 | 1.50 | 1.26 | 1.08 |
| 7739_Personalis | 99'236 | 68 | 96.75 | 12.54 | 5.48 | 1.73 |
| 374_Personalis (additional DNA 1) | 99'236 | 72 | 97.44 | 9.83 | 4.02 | 1.21 |
| 7498_Personalis (additional DNA 2) | 99'236 | 68 | 96.53 | 12.39 | 5.32 | 1.56 |

**Table S12.** Enrichment efficiency of the three enrichment platforms performed by different vendors (V2-V4) for RefSeq exons.

| Agilent_V2_RefSeq_all | 44 | 280 | 326 | 2905 | 7344 | 7739 | Mean |
|---|---|---|---|---|---|---|---|
| Total number of exons | 232'619 | 232'619 | 232'619 | 232'619 | 232'619 | 232'619 | 232'619 |
| Approximate number of reads on exons | 91'956'375 | 85'988'017 | 83'572'762 | 68'364'391 | 82'213'511 | 83'948'932 | 82'673'998 |
| Approximate mean number of reads per exon | 395 | 370 | 359 | 294 | 353 | 361 | 355 |
| Mean read depth | 132 | 121 | 118 | 96 | 116 | 120 | 117 |
| Mean % coverage at 1× | 95.15 | 94.48 | 94.72 | 94.57 | 94.78 | 94.63 | 94.72 |
| Mean % coverage at 20× | 85.93 | 85.79 | 85.77 | 84.83 | 85.65 | 85.58 | 85.59 |
| **NimbleGen_V3_RefSeq_all** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of exons | 232'619 | 232'619 | 232'619 | 232'619 | 232'619 | 232'619 | 232'619 |
| Approximate number of reads on exons | 37'142'009 | 83'642'818 | 31'319'114 | 45'838'652 | 61'584'881 | 20'887'479 | 46'735'825 |
| Approximate mean number of reads per exon | 160 | 360 | 135 | 197 | 265 | 90 | 201 |
| Mean read depth | 48 | 98 | 40 | 56 | 74 | 26 | 57 |
| Mean % coverage at 1× | 95.53 | 96.10 | 95.26 | 95.81 | 96.00 | 94.99 | 95.61 |
| Mean % coverage at 20× | 89.58 | 92.04 | 87.36 | 88.74 | 91.25 | 73.80 | 87.13 |
| **Illumina_V4_RefSeq_all** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of exons | 232'619 | 232'619 | 232'619 | 232'619 | 232'619 | 232'619 | 232'619 |
| Approximate number of reads on exons | 29'121'995 | 24'223'385 | 29'600'776 | 30'262'529 | 29'426'507 | 25'627'386 | 28'043'763 |
| Approximate mean number of reads per exon | 125 | 104 | 127 | 130 | 127 | 110 | 121 |
| Mean read depth | 48 | 41 | 50 | 51 | 48 | 42 | 47 |
| Mean % coverage at 1× | 88.10 | 87.93 | 88.11 | 88.22 | 88.18 | 88.02 | 88.09 |
| Mean % coverage at 20× | 76.25 | 72.23 | 76.57 | 76.74 | 76.39 | 73.18 | 75.23 |
| **Agilent_V2_RefSeq_coding region** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of exons | 202'044 | 202'044 | 202'044 | 202'044 | 202'044 | 202'044 | 202'044 |
| Approximate number of reads on exons | 57'570'213 | 52'867'135 | 51'777'832 | 42'473'247 | 50'455'925 | 52'303'873 | 51'241'371 |
| Approximate mean number of reads per exon | 285 | 262 | 256 | 210 | 250 | 259 | 254 |
| Mean read depth | 156 | 144 | 140 | 114 | 137 | 142 | 139 |
| Mean % coverage at 1× | 99.19 | 99.13 | 99.19 | 99.26 | 99.16 | 99.14 | 99.18 |
| Mean % coverage at 15× | 97.59 | 97.49 | 97.49 | 97.07 | 97.42 | 97.42 | 97.42 |
| Mean % coverage at 20× | 96.96 | 96.82 | 96.81 | 95.92 | 96.69 | 96.67 | 96.65 |
| **NimbleGen_V3_RefSeq_coding region** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of exons | 202'044 | 202'044 | 202'044 | 202'044 | 202'044 | 202'044 | 202'044 |
| Approximate number of reads on exons | 18'364'643 | 39'619'058 | 15'302'804 | 21'727'594 | 29'212'627 | 10'084'786 | 22'385'252 |
| Approximate mean number of reads per exon | 91 | 196 | 76 | 108 | 145 | 50 | 111 |
| Mean read depth | 51 | 104 | 42 | 59 | 79 | 28 | 61 |
| Mean % coverage at 1× | 98.63 | 98.72 | 98.56 | 98.76 | 98.75 | 98.52 | 98.66 |
| Mean % coverage at 15× | 96.71 | 97.65 | 95.78 | 96.26 | 97.34 | 89.99 | 95.62 |
| Mean % coverage at 20× | 95.39 | 97.14 | 93.37 | 94.51 | 96.61 | 79.76 | 92.80 |
| **Illumina_V4_RefSeq_coding region** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of exons | 202'044 | 202'044 | 202'044 | 202'044 | 202'044 | 202'044 | 202'044 |
| Approximate number of reads on exons | 18'228'955 | 15'256'346 | 18'765'557 | 19'068'256 | 18'356'073 | 15'984'498 | 17'609'947 |
| Approximate mean number of reads per exon | 90 | 76 | 93 | 94 | 91 | 79 | 87 |
| Mean read depth | 53 | 45 | 55 | 55 | 52 | 46 | 51 |
| Mean % coverage at 1× | 94.30 | 94.23 | 94.32 | 94.41 | 94.34 | 94.26 | 94.31 |
| Mean % coverage at 15× | 88.09 | 85.39 | 88.24 | 88.33 | 88.27 | 86.20 | 87.42 |
| Mean % coverage at 20× | 83.17 | 79.06 | 83.67 | 83.73 | 83.28 | 79.88 | 82.13 |
| **Agilent_V2_RefSeq_5'UTR** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of exons | 54'367 | 54'367 | 54'367 | 54'367 | 54'367 | 54'367 | 54'367 |
| Approximate number of reads on exons | 9'544'557 | 8'639'899 | 8'605'274 | 7'141'824 | 8'259'030 | 8'543'370 | 8'455'659 |
| Approximate mean number of reads per exon | 176 | 159 | 158 | 131 | 152 | 157 | 156 |
| Mean read depth | 113 | 102 | 101 | 83 | 97 | 101 | 100 |
| Mean % coverage at 1× | 88.33 | 86.25 | 86.87 | 86.31 | 87.06 | 86.78 | 86.93 |
| Mean % coverage at 20× | 62.99 | 62.54 | 62.80 | 61.97 | 62.30 | 62.34 | 62.49 |
| **NimbleGen_V3_RefSeq_5'UTR** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of exons | 54'367 | 54'367 | 54'367 | 54'367 | 54'367 | 54'367 | 54'367 |
| Approximate number of reads on exons | 4'208'508 | 8'995'809 | 3'490'069 | 4'840'571 | 6'547'724 | 2'248'176 | 5'055'143 |
| Approximate mean number of reads per exon | 77 | 165 | 64 | 89 | 120 | 41 | 93 |
| Mean read depth | 39 | 79 | 32 | 44 | 59 | 21 | 46 |
| Mean % coverage at 1× | 91.62 | 92.85 | 90.96 | 91.94 | 92.50 | 90.20 | 91.68 |
| Mean % coverage at 20× | 76.14 | 81.70 | 71.76 | 73.82 | 79.44 | 54.12 | 72.83 |
| **Illumina_V4_RefSeq_5'UTR** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of exons | 54'367 | 54'367 | 54'367 | 54'367 | 54'367 | 54'367 | 54'367 |
| Approximate number of reads on exons | 3'836'677 | 3'138'739 | 3'895'716 | 3'940'321 | 3'835'576 | 3'385'497 | 3'672'088 |
| Approximate mean number of reads per exon | 71 | 58 | 72 | 72 | 71 | 62 | 68 |
| Mean read depth | 39 | 32 | 40 | 40 | 39 | 34 | 37 |
| Mean % coverage at 1× | 75.70 | 75.33 | 75.72 | 75.75 | 75.81 | 75.59 | 75.65 |
| Mean % coverage at 20× | 59.32 | 54.96 | 59.24 | 59.55 | 59.51 | 56.63 | 58.20 |
| **Agilent_V2_RefSeq_3'UTR** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of exons | 38'600 | 38'600 | 38'600 | 38'600 | 38'600 | 38'600 | 38'600 |
| Approximate number of reads on exons | 32'871'014 | 32'041'868 | 30'475'740 | 24'734'719 | 30'719'883 | 30'428'127 | 30'211'892 |
| Approximate mean number of reads per exon | 852 | 830 | 790 | 641 | 796 | 788 | 783 |
| Mean read depth | 111 | 106 | 101 | 82 | 101 | 102 | 100 |
| Mean % coverage at 1× | 89.20 | 88.12 | 88.65 | 88.11 | 88.73 | 88.22 | 88.51 |
| Mean % coverage at 20× | 73.02 | 73.07 | 72.96 | 71.98 | 72.98 | 72.66 | 72.78 |
| **NimbleGen_V3_RefSeq_3'UTR** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of exons | 38'600 | 38'600 | 38'600 | 38'600 | 38'600 | 38'600 | 38'600 |
| Approximate number of reads on exons | 17'649'879 | 41'978'965 | 15'133'918 | 23'142'414 | 30'974'043 | 10'299'582 | 23'196'467 |
| Approximate mean number of reads per exon | 457 | 1088 | 392 | 600 | 802 | 267 | 601 |
| Mean read depth | 45 | 102 | 38 | 56 | 75 | 26 | 57 |
| Mean % coverage at 1× | 89.12 | 90.60 | 88.53 | 89.87 | 90.33 | 87.93 | 89.39 |
| Mean % coverage at 20× | 80.67 | 83.84 | 79.13 | 81.26 | 83.08 | 69.53 | 79.59 |
| **Illumina_V4_RefSeq_3'UTR** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of exons | 38'600 | 38'600 | 38'600 | 38'600 | 38'600 | 38'600 | 38'600 |
| Approximate number of reads on exons | 9'237'507 | 7'648'746 | 9'160'291 | 9'529'750 | 9'436'549 | 8'171'583 | 8'864'071 |
| Approximate mean number of reads per exon | 239 | 198 | 237 | 247 | 244 | 212 | 230 |
| Mean read depth | 31 | 26 | 31 | 32 | 31 | 27 | 30 |
| Mean % coverage at 1× | 75.45 | 74.74 | 75.34 | 75.63 | 75.72 | 75.16 | 75.34 |
| Mean % coverage at 20× | 54.27 | 48.44 | 53.00 | 54.55 | 55.19 | 50.81 | 52.71 |

**Table S13.** Enrichment efficiency of the three enrichment platforms performed by different vendors (V2-V4) for intronic sequences (regions) in our region of interest (RefSeq coding exons and -50-bp and +20-bp flanking intronic sequences).

| Agilent_V2_RefSeq_coding_-50 bp | 44 | 280 | 326 | 2905 | 7344 | 7739 | Mean |
|---|---|---|---|---|---|---|---|
| Total number of regions | 199'107 | 199'107 | 199'107 | 199'107 | 199'107 | 199'107 | **199'107** |
| Approximate number of reads on regions | 11'745'824 | 10'987'395 | 10'673'000 | 8'776'634 | 10'526'929 | 10'697'536 | **10'567'886** |
| Approximate mean number of reads per region | 59 | 55 | 54 | 44 | 53 | 54 | **53** |
| Mean read depth | 118 | 110 | 107 | 88 | 106 | 107 | **106** |
| Mean % coverage at 15× | 95.36 | 95.59 | 95.49 | 94.14 | 95.51 | 94.97 | **95.18** |
| Mean % coverage at 20× | 93.20 | 93.58 | 93.35 | 90.96 | 93.40 | 92.60 | **92.85** |
| **NimbleGen_V3_RefSeq_coding_-50 bp** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of regions | 199'107 | 199'107 | 199'107 | 199'107 | 199'107 | 199'107 | **199'107** |
| Approximate number of reads on regions | 4'193'288 | 8'643'131 | 3'535'739 | 4'957'626 | 6'606'299 | 2'341'553 | **5'046'273** |
| Approximate mean number of reads per region | 21 | 43 | 18 | 25 | 33 | 12 | **25** |
| Mean read depth | 42 | 87 | 36 | 50 | 66 | 24 | **51** |
| Mean % coverage at 15× | 95.04 | 97.23 | 93.01 | 94.58 | 96.68 | 81.11 | **92.94** |
| Mean % coverage at 20× | 91.34 | 96.36 | 86.93 | 91.25 | 95.34 | 64.00 | **87.54** |
| **Illumina_V4_RefSeq_coding_-50 bp** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of regions | 199'107 | 199'107 | 199'107 | 199'107 | 199'107 | 199'107 | **199'107** |
| Approximate number of reads on regions | 2'636'469 | 2'207'057 | 2'618'019 | 2'748'245 | 2'670'685 | 2'324'392 | **2'534'145** |
| Approximate mean number of reads per region | 13 | 11 | 13 | 14 | 13 | 12 | **13** |
| Mean read depth | 26 | 22 | 26 | 28 | 27 | 23 | **25** |
| Mean % coverage at 15× | 59.01 | 52.53 | 58.01 | 60.18 | 59.84 | 54.55 | **57.35** |
| Mean % coverage at 20× | 47.52 | 41.00 | 46.69 | 49.09 | 48.31 | 42.68 | **45.88** |
| **Agilent_V2_RefSeq_coding_+20 bp** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of regions | 197'564 | 197'564 | 197'564 | 197'564 | 197'564 | 197'564 | **197'564** |
| Approximate number of reads on regions | 5'151'344 | 4'799'004 | 4'665'424 | 3'820'170 | 4'585'481 | 4'691'331 | **4'618'792** |
| Approximate mean number of reads per region | 26 | 24 | 24 | 19 | 23 | 24 | **23** |
| Mean read depth | 130 | 121 | 118 | 97 | 116 | 119 | **117** |
| Mean % coverage at 15× | 96.39 | 96.42 | 96.39 | 95.43 | 96.31 | 96.11 | **96.17** |
| Mean % coverage at 20× | 94.90 | 94.89 | 94.84 | 93.11 | 94.70 | 94.36 | **94.47** |
| **NimbleGen_V3_RefSeq_coding_+20 bp** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of regions | 197'564 | 197'564 | 197'564 | 197'564 | 197'564 | 197'564 | **197'564** |
| Approximate number of reads on regions | 1'841'585 | 3'836'958 | 1'553'456 | 2'177'374 | 2'914'931 | 1'021'349 | **2'224'276** |
| Approximate mean number of reads per region | 9 | 19 | 8 | 11 | 15 | 5 | **11** |
| Mean read depth | 47 | 97 | 39 | 55 | 74 | 26 | **56** |
| Mean % coverage at 15× | 96.11 | 97.42 | 94.77 | 95.52 | 97.03 | 86.34 | **94.53** |
| Mean % coverage at 20× | 93.99 | 96.79 | 91.08 | 93.09 | 96.08 | 72.77 | **90.63** |
| **Illumina_V4_RefSeq_coding_+20 bp** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of regions | 197'564 | 197'564 | 197'564 | 197'564 | 197'564 | 197'564 | **197'564** |
| Approximate number of reads on regions | 1'270'409 | 1'059'550 | 1'275'194 | 1'325'726 | 1'287'699 | 1'119'616 | **1'223'032** |
| Approximate mean number of reads per region | 6 | 5 | 6 | 7 | 7 | 6 | **6** |
| Mean read depth | 32 | 27 | 32 | 34 | 33 | 28 | **31** |
| Mean % coverage at 15× | 72.12 | 65.46 | 71.51 | 72.77 | 72.80 | 68.02 | **70.45** |
| Mean % coverage at 20× | 61.06 | 53.39 | 60.71 | 62.41 | 61.81 | 55.83 | **59.20** |

**Table S14.** Enrichment efficiency of the three platforms performed by the same vendor (V1) for our panel of eight genes (exons including UTR).

| Agilent_V1_gene panel | 44 | 280 | 326 | 2905 | 7344 | 7739 | Mean |
|---|---|---|---|---|---|---|---|
| Total number of exons | 169 | 169 | 169 | 169 | 169 | 169 | **169** |
| Approximate number of reads on exons | 43'286 | 47'277 | 35'254 | 39'553 | 53'291 | 49'016 | **44'612** |
| Approximate mean number of reads per exon | 256 | 280 | 209 | 234 | 315 | 290 | **264** |
| Mean read depth | 95 | 103 | 76 | 85 | 117 | 108 | **97** |
| Mean % coverage at 1× | 97.54 | 97.54 | 97.43 | 97.62 | 97.93 | 98.28 | **97.72** |
| Mean % coverage at 20× | 93.85 | 95.29 | 91.75 | 93.18 | 94.96 | 94.43 | **93.91** |
| **NimbleGen_V1_gene panel** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of exons | 169 | 169 | 169 | 169 | 169 | 169 | **169** |
| Approximate number of reads on exons | 66'968 | 48'757 | 47'518 | 55'857 | 49'783 | 60'856 | **54'956** |
| Approximate mean number of reads per exon | 396 | 289 | 281 | 331 | 295 | 360 | **325** |
| Mean read depth | 133 | 92 | 91 | 110 | 99 | 124 | **108** |
| Mean % coverage at 1× | 96.36 | 94.48 | 96.11 | 96.51 | 96.42 | 96.09 | **96.00** |
| Mean % coverage at 20× | 92.48 | 85.32 | 88.16 | 90.46 | 91.05 | 91.32 | **89.80** |
| **Illumina_V1_gene panel** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of exons | 169 | 169 | 169 | 169 | 169 | 169 | **169** |
| Approximate number of reads on exons | 26'343 | 29'196 | 25'726 | 32'869 | 29'335 | 23'761 | **27'872** |
| Approximate mean number of reads per exon | 156 | 173 | 152 | 194 | 174 | 141 | **165** |
| Mean read depth | 73 | 80 | 74 | 90 | 79 | 66 | **77** |
| Mean % coverage at 1× | 98.63 | 98.87 | 98.61 | 98.91 | 98.91 | 98.80 | **98.79** |
| Mean % coverage at 20× | 89.48 | 91.22 | 88.69 | 89.90 | 91.03 | 87.13 | **89.58** |

**Table S15.** Enrichment efficiency of the three platforms performed by different vendors (V2-V4) for our panel of eight genes (exons including UTR).

| Agilent_V2_gene panel | 44 | 280 | 326 | 2905 | 7344 | 7739 | Mean |
|---|---|---|---|---|---|---|---|
| Total number of exons | 169 | 169 | 169 | 169 | 169 | 169 | **169** |
| Approximate number of reads on exons | 62'943 | 60'283 | 58'613 | 47'919 | 58'348 | 59'046 | **57'859** |
| Approximate mean number of reads per exon | 372 | 357 | 347 | 284 | 345 | 349 | **342** |
| Mean read depth | 140 | 130 | 128 | 105 | 127 | 130 | **127** |
| Mean % coverage at 1× | 98.81 | 99.35 | 99.07 | 98.40 | 99.23 | 98.91 | **98.96** |
| Mean % coverage at 20× | 95.77 | 96.54 | 95.98 | 95.12 | 96.38 | 95.82 | **95.94** |
| **NimbleGen_V3_gene panel** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of exons | 169 | 169 | 169 | 169 | 169 | 169 | **169** |
| Approximate number of reads on exons | 27'184 | 62'876 | 23'432 | 35'907 | 46'672 | 16'157 | **35'371** |
| Approximate mean number of reads per exon | 161 | 372 | 139 | 212 | 276 | 96 | **209** |
| Mean read depth | 53 | 115 | 45 | 66 | 86 | 31 | **66** |
| Mean % coverage at 1× | 99.99 | 100.00 | 100.00 | 100.00 | 100.00 | 99.79 | **99.96** |
| Mean % coverage at 20× | 96.42 | 98.53 | 96.31 | 96.51 | 98.63 | 88.86 | **95.88** |
| **Illumina_V4_gene panel** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Mean** |
| Total number of exons | 169 | 169 | 169 | 169 | 169 | 169 | **169** |
| Approximate number of reads on exons | 21'118 | 17'873 | 22'377 | 22'273 | 21'260 | 19'030 | **20'655** |
| Approximate mean number of reads per exon | 125 | 106 | 132 | 132 | 126 | 113 | **122** |
| Mean read depth | 59 | 51 | 64 | 63 | 59 | 52 | **58** |
| Mean % coverage at 1× | 98.93 | 98.73 | 98.77 | 98.71 | 98.81 | 98.52 | **98.74** |
| Mean % coverage at 20× | 90.59 | 89.01 | 91.84 | 93.09 | 91.21 | 89.76 | **90.92** |

**Table S16.** Number of heterozygous SNVs and indels identified by Sanger sequencing in the six DNA samples of this study. The number of homozygous SNVs is given in parentheses.

| Sample | SNVs | | | | Indels | | | |
|---|---|---|---|---|---|---|---|---|
| | Total | ROI | Coding | UTR | Total | ROI | Coding | UTR |
| 44 | 16 (2) | 4 (1) | 1 (0) | 2 (1) | 3 (1) | 0 (0) | 0 (0) | 0 (0) |
| 280 | 10 (0) | 4 (0) | 0 (0) | 1 (0) | 3 (0) | 1 (0) | 1 (0) | 0 (0) |
| 326 | 1 (2) | 1 (1) | 0 (0) | 0 (0) | 1 (0) | 1 (0) | 1 (0) | 0 (0) |
| 2905 | 42 (2) | 16 (2) | 6 (0) | 1 (0) | 6 (0) | 2 (0) | 0 (0) | 0 (0) |
| 7344 | 39 (2) | 18 (1) | 8 (0) | 1 (1) | 6 (1) | 3 (0) | 1 (0) | 0 (0) |
| 7739 | 17 (3) | 6 (0) | 2 (0) | 2 (0) | 1 (0) | 0 (0) | 0 (0) | 0 (0) |
| **Total** | **125 (11)** | **49 (5)** | **17 (0)** | **7 (2)** | **20 (2)** | **7 (0)** | **3 (0)** | **0 (0)** |

ROI: region of interest (coding region with 50-bp upstream and 20-bp downstream intronic sequences).

**Table S17.** Efficiency and reproducibility of the three enrichment platforms for different heterozygous SNVs detected by Sanger sequencing. In the number pairs (m/n), the first number (m) represents number of different heterozygous SNVs obtained in at least one sample with sufficient number of reads (≥20). The second number (n) represents number of heterozygous positions detected by Sanger sequencing, which were identified in particular data sets with correct genotype and by at least one read.

| Total | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Total SNVs | | | | | | SNVs in designed target region | | | | | | SNVs in designed target region of all three platforms | | | | | |
| Heterozygous | AG (V1) | AG (V2) | NG (V1) | NG (V3) | ILL (V1) | ILL (V4) | AG (V1) | AG (V2) | NG (V1) | NG (V3) | ILL (V1) | ILL (V4) | AG (V1) | AG (V2) | NG (V1) | NG (V3) | ILL (V1) | ILL (V4) |
| 1/6 samples | 30/44 | 36/43 | 39/42 | 38/44 | 28/42 | 21/32 | 21/26 | 24/26 | 30/30 | 30/31 | 9/9 | 8/9 | 8/8 | 8/8 | 8/8 | 8/8 | 8/8 | 7/8 |
| 2/6 samples | 21/24 | 22/25 | 23/24 | 22/25 | 12/24 | 10/16 | 13/13 | 13/13 | 17/17 | 16/17 | 6/6 | 6/6 | 6/6 | 6/6 | 6/6 | 6/6 | 6/6 | 6/6 |
| 3/6 samples | 7/7 | 7/7 | 7/7 | 5/6 | 1/5 | 1/5 | 3/3 | 3/3 | 6/6 | 5/6 | 1/1 | 1/1 | 1/1 | 1/1 | 1/1 | 1/1 | 1/1 | 1/1 |
| 4/6 samples | 2/2 | 2/2 | 2/2 | 0/1 | 0/2 | 0 | 1/1 | 1/1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *>1 sample* | *30/33* | *31/34* | *32/33* | *27/32* | *13/31* | *11/21* | *17/17* | *17/17* | *23/23* | *21/23* | *7/7* | *7/7* | *7/7* | *7/7* | *7/7* | *7/7* | *7/7* | *7/7* |
| Positions with CR | 2 | 1 | 1 | 4 | 2 | 5 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| True results | 77 | 77 | 75 | 76 | 73 | 53 | 43 | 43 | 53 | 54 | 16 | 16 | 15 | 15 | 15 | 15 | 15 | 15 |
| False and NA results | 1 | 1 | 3 | 2 | 5 | 25 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Region of interest (ROI, coding region with 50-bp upstream and 20-bp downstream intronic sequences)** | | | | | | | | | | | | | | | | | | |
| | Total SNVs | | | | | | SNVs in designed target region | | | | | | SNVs in designed target region of all three platforms | | | | | |
| Heterozygous | AG (V1) | AG (V2) | NG (V1) | NG (V3) | ILL (V1) | ILL (V4) | AG (V1) | AG (V2) | NG (V1) | NG (V3) | ILL (V1) | ILL (V4) | AG (V1) | AG (V2) | NG (V1) | NG (V3) | ILL (V1) | ILL (V4) |
| 1/6 samples | 13/16 | 14/16 | 16/16 | 16/16 | 14/16 | 14/16 | 12/15 | 13/15 | 16/16 | 16/16 | 6/6 | 6/6 | 6/6 | 6/6 | 6/6 | 6/6 | 6/6 | 6/6 |
| 2/6 samples | 9/9 | 9/9 | 9/9 | 8/9 | 7/9 | 8/9 | 9/9 | 9/9 | 9/9 | 8/9 | 4/4 | 4/4 | 4/4 | 4/4 | 4/4 | 4/4 | 4/4 | 4/4 |
| 3/6 samples | 5/5 | 5/5 | 5/5 | 4/5 | 1/5 | 1/5 | 2/2 | 2/2 | 5/5 | 4/5 | 1/1 | 1/1 | 1/1 | 1/1 | 1/1 | 1/1 | 1/1 | 1/1 |
| 4/6 samples | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *>1 sample* | 14/14 | 14/14 | 14/14 | 12/14 | 8/14 | 9/14 | 11/11 | 11/11 | 14/14 | 12/14 | 5/5 | 5/5 | 5/5 | 5/5 | 5/5 | 5/5 | 5/5 | 5/5 |
| Positions with CR | 0 | 0 | 0 | 2 | 1 | 3 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| True results | 30 | 30 | 30 | 30 | 30 | 30 | 26 | 26 | 30 | 30 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 |
| False and NA results | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Untranslated regions (UTR)** | | | | | | | | | | | | | | | | | | |
| | Total SNVs | | | | | | SNVs in designed target region | | | | | | SNVs in designed target region of all three platforms | | | | | |
| Heterozygous | AG (V1) | AG (V2) | NG (V1) | NG (V3) | ILL (V1) | ILL (V4) | AG (V1) | AG (V2) | NG (V1) | NG (V3) | ILL (V1) | ILL (V4) | AG (V1) | AG (V2) | NG (V1) | NG (V3) | ILL (V1) | ILL (V4) |
| 1/6 samples | 2/3 | 2/3 | 2/2 | 2/3 | 3/3 | 2/3 | 2/2 | 2/2 | 2/2 | 2/3 | 3/3 | 2/3 | 2/2 | 2/2 | 2/2 | 2/2 | 2/2 | 1/2 |
| 2/6 samples | 2/2 | 2/2 | 2/2 | 2/2 | 2/2 | 2/2 | 2/2 | 2/2 | 2/2 | 2/2 | 2/2 | 2/2 | 2/2 | 2/2 | 2/2 | 2/2 | 2/2 | 2/2 |
| 3/6 samples | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4/6 samples | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *>1 sample* | 2/2 | 2/2 | 2/2 | 2/2 | 2/2 | 2/2 | 2/2 | 2/2 | 2/2 | 2/2 | 2/2 | 2/2 | 2/2 | 2/2 | 2/2 | 2/2 | 2/2 | 2/2 |
| Positions with CR | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| True results | 5 | 5 | 4 | 5 | 5 | 5 | 4 | 4 | 4 | 5 | 5 | 5 | 4 | 4 | 4 | 4 | 4 | 4 |
| False and NA results | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

SNV: single nucleotide variant; AG: Agilent; NG: NimbleGen; ILL: Illumina; CR: conflicting results, indicating number of heterozygous variants with ≥20 reads in at least one sample, for which in other sample(s) this read depth was not reached; NA: not applicable (here: zero reads).

**Table S18.** Heterozygous SNVs identified by Sanger sequencing but not detected by at least one of the three WES platforms. SNVs are specified by gene, gDNA position, number of samples, enrichment platform (vendor) with false result, and read depth. In the right half of the table, true results at the same positions obtained in remaining data sets for the same sample(s) are presented.

| Gene | DNA position | Within ROI | No. of samples | Platform | Design | Result (no. of results /no. of samples) | Change | NOR (sample) | Platform | Design | Result (no. of results /no. of samples) | NOR (sample) | Other info |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *COL3A1* | Chr2:189857993T>C | no | 2 | ILL (V4) | flanking | false positive (2/2) | P--->PH | 1 (2905), 2 (7344) | AG (V2) | flanking | true (2/2) | 16 (2905), 28 (7344) | OK with V1; AG (V2 and V1): CR (2905) |
| | | | | | | | | | NG (V3) | flanking | true (2/2) | 67 (2905), 56 (7344) | |
| | Chr2:189860630A>C | no | 1 | ILL (V4) | --- | NA (1/1) | P--->NA | 0 (2905) | AG (V2) | flanking | true (1/1) | 20 (2905) | OK with V1; AG (V1) NOR<20 |
| | | | | | | | | | NG (V3) | flanking | true(1/1) | 28 (2905) | |
| | Chr2:189862352G>A | no | 1 | ILL (V4) | flanking | NA (1/1) | P--->NA | 0 (2905) | AG (V2) | in design | true (1/1) | 60 (2905) | OK with V1; ILL (V1) NOR<20 |
| | | | | | | | | | NG (V3) | in design | true(1/1) | 52 (2905) | |
| | Chr2:189862890G>A | no | 1 | ILL (V4) | --- | NA (1/1) | P--->NA | 0 (2905) | AG (V2) | flanking | true (1/1) | 33 (2905) | OK with V1; ILL (V1) NOR<20 |
| | | | | | | | | | NG (V3) | flanking | true(1/1) | 33 (2905) | |
| | Chr2:189867882C>T | no | 1 | ILL (V4) | flanking | false positive (1/1) | P--->PH | 1 (7344) | AG (V2) | flanking | true (1/1) | 22 (7344) | OK with V1; AG (V1): NOR<20 |
| | | | | | | | | | NG (V3) | --- | true(1/1) | 78 (7344) | |
| | Chr2:189871573A>G | no | 1 | ILL (V4) | flanking | false negative (1/1) | P--->WT | 1 (2905) | AG (V2) | flanking | true (1/1) | 16 (2905) | OK with V1; AG (V1) and ILL (V1) NOR<20 |
| | | | | | | | | | NG (V3) | in design | true(1/1) | 45 (2905) | |
| | Chr2:189875237T>G | no | 1 | ILL (V4) | --- | NA (1/1) | P--->NA | 0 (2905) | AG (V2) | --- | true (1/1) | 30 (2905) | OK with V1; AG (V1) NOR<20 |
| | | | | | | | | | NG (V3) | flanking | true(1/1) | 74 (2905) | |
| *FBN1* | Chr15:48903126A>G | no | 4 | ILL (V4) | --- | NA (4/4) | P--->NA | 0 (44), 0 (280), 0 (7344), 0 (7739) | AG (V2) | in design | true (4/4) | 113 (44), 96 (280), 80 (7344), 78 (7739) | OK with V1; ILL NOR<20; NG (V3): CR |
| | | | | | | | | | NG (V3) | flanking | true (4/4) | 13 (44), 50 (280), 42 (7344), 15 (7739) | |
| | Chr15:48813107G>T | no | 2 | ILL (V4) | flanking | NA (1/2) | P--->NA | 0 (2905) | AG (V2) | flanking | true (2/2) | 48 (2905), 31 (7344) | OK with V1; ILL (V1) NOR<20; NG (V1 and V3): CR |
| | | | | | | | | | NG (V3) | flanking | true (2/2) | 14 (2905), 27 (7344) | |
| | | | | | | | | | ILL (V4) | flanking | true (1/2) | 2 (7344) | |
| | Chr15:48797424C>A | no | 2 | ILL (V4) | flanking | false positive (1/2) | P--->PH | 3 (7344) | AG (V2) | flanking | true (2/2) | 22 (2905), 31 (7344) | OK with V1; AG (V1): CR; ILL (V1) NOR<20 |
| | | | | | | | | | NG (V3) | flanking | true (2/2) | 31 (2905), 49 (7344) | |
| | | | | | | | | | ILL (V4) | flanking | true (1/2) | 7 (7344) | |
| | Chr15:48797124T>A | no | 2 | ILL (V4) | flanking | NA (2/2) | P--->NA | 0 (44), 0 (2905) | AG (V2) | flanking | true (2/2) | 51 (44), 35 (2905) | OK with V1; ILL (V1) NOR<20 |
| | | | | | | | | | NG (V3) | flanking | true (2/2) | 40 (44), 47 (2905) | |
| | Chr15:48795877G>A | no | 2 | ILL (V4) | --- | NA (2/2) | P--->NA | 0 (2905), 0 (7344) | AG (V2) | flanking | true (2/2) | 9 (2905), 8 (7344) | |
| | | | | | | | | | NG (V3) | flanking | true (2/2) | 20 (44), 34 (7344) | |
| | | | | AG (V1) | flanking | NA (1/2) | P--->NA | 0 (7344) | NG (V1) | flanking | true (2/2) | 40 (2905), 41 (7344) | |
| | | | | | | | | | ILL (V1) | --- | true (2/2) | 15 (2905), 9 (7344) | |
| | Chr15:48784533G>A | no | 1 | ILL (V4) | --- | NA (1/1) | P--->NA | 0 (7739) | AG (V2) | --- | true (1/1) | 16 (7739) | OK with V1; all NOR<20 |
| | | | | | | | | | NG (V3) | flanking | true (1/1) | 6 (7739) | |
| | Chr15:48779138G>A | no | 2 | ILL (V4) | --- | NA (2/2) | P--->NA | 0 (44), 0 (7344) | AG (V2) | --- | true (2/2) | 10 (44), 3 (7344) | |
| | | | | | | | | | NG (V3) | --- | true (2/2) | 10 (44), 15 (7344) | |
| | | | | NG (V1) | --- | false negative (1/2) | P--->WT | 4 (44) | AG (V1) | --- | true (2/2) | 6 (44), 3 (7344) | |
| | | | | | | | | | NG (V1) | --- | true (1/2) | 2 (7344) | |
| | | | | ILL (V1) | --- | false negative (1/2) | P--->WT | 2 (44) | ILL (V1) | --- | true (1/2) | 2 (7344) | |
| | Chr15:48760391G>A | no | 2 | ILL (V4)* | flanking | false negative (1/2) | P--->WT | 1 (2905) | AG (V2) | flanking | true (2/2) | 34 (2905), 37 (7344) | OK with V1; ILL (V1) NOR<20 |
| | | | | ILL (V4)* | flanking | false positive (1/2) | P--->PH | 2 (7344) | NG (V3) | flanking | true (2/2) | 37 (2905), 46 (7344) | |
| | Chr15:48758132A>C | no | 4 | NG (V3) | flanking | false positive (1/4) | P--->PH | 10 (7739) | AG (V2) | flanking | true (4/4) | 43 (44), 45 (280), 50 (7344), 46 (7739) | OK with V1; ILL (V1): CR; NG (V3): also CR |
| | | | | | | | | | NG (V3) | flanking | true (3/4) | 15 (44), 61 (280), 27 (7344) | |
| | | | | ILL (V4) | flanking | false positive (1/4) | P--->PH | 5 (7344) | ILL (V4) | flanking | true (3/4) | 4 (44), 3 (280), 11 (7739) | |
| | Chr15:48736684A>T | no | 3 | ILL (V4) | flanking | false negative (1/3) | P--->WT | 11 (280) | AG (V2) | in design | true (3/3) | 137 (280), 120 (2905), 157 (7739) | |
| | | | | | | | | | NG (V3) | in design | true (3/3) | 104 (280), 55 (2905), 31 (7739) | |
| | | | | | | | | | ILL (V4) | flanking | true (2/3) | 7 (2905), 4 (7739) | |
| | | | | ILL (V1) | flanking | false negative (1/3) | P--->WT | 1 (7739) | AG (V1) | in design | true (3/3) | 148 (280), 103 (2905), 133 (7739) | |
| | | | | | | | | | NG (V1) | in design | true (3/3) | 102 (280), 100 (2905), 123 (7739) | |
| | | | | | | | | | ILL (V1) | flanking | true (2/3) | 6 (280), 6 (2905) | |
| | Chr15:48736615T>C | no | 2 | ILL (V4) | --- | NA (2/2) | P--->NA | 0 (7344), 0 (7739) | AG (V2) | flanking | true (2/2) | 51 (7344), 60 (7739) | OK with V1; ILL NOR<20 |
| | | | | | | | | | NG (V3) | flanking | true (2/2) | 30 (7344), 41 (7739) | |
| | Chr15:48729359A>T | no | 2 | ILL (V4) | flanking | false negative (1/2) | P--->WT | 6 (2905) | AG (V2) | flanking | true (2/2) | 61 (2905), 96 (7344) | OK with V1; ILL NOR<20 |
| | | | | | | | | | NG (V3) | in design | true (2/2) | 53 (2905), 75 (7344) | |
| | | | | | | | | | ILL (V4) | flanking | true (1/2) | 5 (7344) | |
| | Chr15:48729030T>C | no | 3 | NG (V3) | flanking | false negative (1/3) | P--->WT | 5 (7739) | AG (V2) | --- | true (3/3) | 54 (280), 31 (7344), 29 (7739) | |
| | | | | ILL (V4) | --- | NA (3/3) | P--->NA | 0 (280), 0 (7344), 0 (7739) | NG (V3) | flanking | true (2/3) | 43 (280), 30 (7344) | |
| | | | | | | | | | AG (V1) | --- | true (3/3) | 54 (280), 31 (7344), 29 (7739) | |
| | | | | ILL (V1) | --- | false positive (1/3) | P--->PH | 1 (7739) | NG (V1) | flanking | true (2/3) | 43 (280), 30 (7344) | |
| | | | | | | | | | ILL (V1) | --- | true (2/3) | 54 (280), 31 (7344), 29 (7739) | |
| | Chr15:48713996G>T | no | 1 | ILL (V4) | --- | NA (1/1) | P--->NA | 0 (44) | AG (V2) | flanking | true (1/1) | 23 (44) | OK with V1; ILL NOR<20 |
| | | | | | | | | | NG (V3) | flanking | true (1/1) | 33 (44) | |
| *SMAD3* | Chr15:67358478G>A | UTR | 1 | NG (V1) | in design | NA (1/1) | P--->NA | 0 (7739) | AG (V1) | flanking | true (1/1) | 14 (7739) | AG (V2) and NG (V3) NOR<20, ILL (V4) NOR >20 |
| | | | | | | | | | ILL (V1) | in design | true (1/1) | 26 (7739) | |
| | Chr15:67457485G>C | no | 1 | NG (V1) | flanking | false positive (1/1) | P--->PH | 1 (7739) | AG (V1) | flanking | true (1/1) | 63 (7739) | AG (V2) and NG (V3) NOR<20, ILL (V4) NOR >20 |
| | | | | | | | | | ILL (V1) | flanking | true (1/1) | 11 (7739) | |
| | Chr15:67458930C>T | no | 1 | ILL (V4) | --- | NA (1/1) | P--->NA | 0 (7739) | AG (V2) | flanking | true (1/1) | 117 (7739) | OK with V1; ILL NOR<20 |
| | | | | | | | | | NG (V3) | --- | true (1/1) | 7 (7739) | |
| | Chr15:67459013C>T | no | 1 | ILL (V4) | --- | NA (1/1) | P--->NA | 0 (7739) | AG (V2) | flanking | true (1/1) | 122 (7739) | OK with V1; all NOR>20 |
| | | | | | | | | | NG (V3) | flanking | true (1/1) | 15 (7739) | |
| *TGFBR1* | Chr9:101907279A>G | no | 1 | ILL (V4) | --- | NA (1/1) | P--->NA | 0 (44) | AG (V2) | --- | true (1/1) | 13 (44) | |
| | | | | | | | | | NG (V3) | flanking | true (1/1) | 15 (44) | |
| | | | | ILL (V1) | --- | NA (1/1) | P--->NA | 0 (44) | AG (V1) | --- | true (1/1) | 10 (44) | |
| | | | | | | | | | NG (V1) | flanking | true (1/1) | 34 (44) | |
| | Chr9:101911321A>G | no | 1 | AG (V2) | --- | false positive (1/1) | P--->PH | 9 (44) | NG (V3) | --- | true (1/1) | 6 (44) | |
| | | | | ILL (V4) | --- | NA (1/1) | P--->NA | 0 (44) | AG (V1) | --- | true (1/1) | 3 (44) | |
| | | | | ILL (V1) | --- | false positive (1/1) | P--->PH | 1 (44) | NG (V1) | --- | true (1/1) | 2 (44) | |

*In Supplementary Table S17 counted only once (i.e., one false result) due to same platform and vendor; NOR: Number of reads; AG: Agilent; NG: NimbleGen; ILL: Illumina; flanking: within 100 bp of design; OK, heterozygous position detected; CR: conflicting results, indicating number of heterozygous variants with ≥20 reads in at least one sample, for which in other sample(s) this read depth was not reached; design: designed target region; WT: wild-type; P: heterozygous SNV; PH: homozygous SNV; NA: not applicable (here: zero reads).

**Table S19.** False positive VCF calls in our ROI and UTR. Variants called heterozygous by WES, but wild-type in Sanger sequencing.

| Sample | Gene | Chromosome | gDNA position | cDNA position | Location | WT allele | Mutant allele | Change | Platform | Vendor |
|---|---|---|---|---|---|---|---|---|---|---|
| 44 | *FBN1* | 15 | 48'760'120 | c.4747+15 | intron 37 | A (7) | T (2) | WT → P | Illumina | V4 |
| 7739 | *FBN1* | 15 | 48'718'056 | c.7210 | exon 58 | C (16) | A (2) | WT → P | Illumina | V4 |
| 7739 | *TGFBR2* | 3 | 30'648'342 | c.-34 | 5'UTR | G (5) | T (2) | WT → P | Illumina | V4 |

WT: wild-type; P: heterozygous SNV; calls in unfiltered VCF provided by vendors (cf. Supplementary Table S3).

**Table S20.** Heterozygous indels in our gene panel detected by Sanger sequencing in the six DNA samples of this study. The values are given as called by the Integrative Genomics Viewer (IGV): mean read depth across the variant (excluding reads carrying a deletion)/reads carrying the variant. Analysis was performed using unfiltered VCF files provided by vendors (cf. Supplementary Table S3).

| Sample | Gene | Heterozygous indels detected by Sanger | Agilent_V1 | NimbleGen_V1 | Illumina_V1 | Agilent_V2 | NimbleGen_V3 | Illumina_V4 |
|---|---|---|---|---|---|---|---|---|
| 44 | *FBN1* | c.3209-72_67delTCTTTA | ***106/21***[1] | ***115/25***[1] | **16/5**[2] | ***141/58***[1] | *58/0*[1] | **15/4**[1] |
| | *FBN1* | c.3589+36_40delTTTTA | **29/6**[2] | 3/3[2] | **20/8**[2] | **46/10**[1] | 23/0[2] | **16/4**[2] |
| | *FBN1* | c.3589+63_67delTTATG (homozygous) | 17/5[2] | 3/2[2] | 5/1[2] | 24/5[1] | 14/0[2] | 0/1[2] |
| | *TGFBR1* | c.1386+90_94delTCTTT | **26/9**[2] | **23/12**[2] | 1/1[2] | **34/20**[1] | **16/15**[1] | NA[2] |
| 280 | *FBN1* | c.3209-72_67delTCTTTA | ***106/26***[1] | ***57/21***[1] | 12/3[2] | ***110/58***[1] | *99/0*[1] | 20/2[1] |
| | *FBN1* | c.3589+63_67delTTATG | 33/3[2] | 2/0[2] | 15/2[2] | 23/20[1] | 40/1[2] | 4/1[2] |
| | *FBN1* | c.5924_5925dupAT (p.E1976MfsX5)* | ***85/28***[1] | ***86/29***[1] | ***24/11***[1] | ***93/55***[1] | ***107/31***[1] | ***20/5***[1] |
| 326 | *FBN1* | c.1904_1919del (p.Y635SfsX78)* | ***37/7***[2] | ***95/12***[2] | *37/3*[2] | ***55/30***[1] | *60/0*[1] | ***33/6***[1] |
| 2905 | *COL3A1* | c.2391+28delC | **13/14**[2] | ***96/54***[1] | **24/15**[2] | **30/13**[1] | **58/32**[1] | **16/14**[1] |
| | *FBN1* | c.3589+36_40delTTTTA | **27/11**[2] | 7/1[2] | **21/5**[2] | **26/11**[1] | 33/0[2] | **28/5**[2] |
| | *FBN1* | c.3589+63_67delTTATG | 18/3[2] | 3/1[2] | 9/1[2] | 20/3[1] | 19/0[2] | 4/0[2] |
| | *FBN1* | c.5066-14dupT** | **119/42**[2] | ***120/57***[1] | **19/5**[2] | **166/74**[1] | ***67/19***[1] | **30/9**[2] |
| | *FBN1* | c.5423-30_28delCCT** | **25/12**[2] | ***24/11***[1] | **59/26**[2] | **23/24**[1] | ***29/17***[1] | **43/15**[1] |
| | *FBN1* | c.5671+28dupT | **67/24**[2] | ***26/12***[1] | **21/11**[2] | **67/28**[1] | ***34/20***[1] | **12/4**[2] |
| 7344 | *FBN1* | c.3209-72_67delTCTTTA | ***123/28***[1] | ***85/25***[1] | **13/6**[2] | ***136/58***[1] | *69/0*[1] | **19/7**[1] |
| | *FBN1* | c.3589+36_40delTTTTA | **26/9**[2] | **7/7**[2] | 17/0[2] | **30/17**[1] | 41/0[2] | 15/1[2] |
| | *FBN1* | c.3589+63_67delTTATG (homozygous) | 15/2[2] | 5/1[2] | 6/3[2] | 21/1[1] | 22/0[2] | 2/4[2] |
| | *FBN1* | c.5066-14dupT** | **150/60**[2] | ***98/37***[1] | 18/1[2] | **165/63**[1] | ***93/25***[1] | **36/16**[2] |
| | *FBN1* | c.5423-30_28delCCT** | **33/21**[2] | ***16/14***[1] | **46/25**[2] | **39/32**[1] | ***21/31***[1] | **37/14**[1] |
| | *FBN1* | c.5671+28dupT | **88/35**[2] | ***20/5***[2] | **14/6**[2] | **88/40**[1] | ***38/17***[1] | **13/9**[2] |
| | *TGFBR1* | c.70_78del (p.A24_A26del)* | 1/0[2] | *NA*[2] | NA[2] | 5/0[2] | *4/0*[2] | NA[2] |
| 7739 | *FBN1* | c.3589+63_67delTTATG | 14/1[2] | 1/0[2] | 7/0[2] | **9/4**[1] | 9/0[2] | 3/3[2] |
| Summary IGV total (unambiguous calls/total indels) | | | 16/22 | 14/22 | 11/22 | 18/22 | 9/22 | 13/22 |
| Summary VCF total (reported in VCF/total indels) | | | 4/22*** | 10/22*** | 1/22*** | 21/22 | 13/22 | 8/22 |
| **Summary IGV ROI (unambiguous calls/total indels)** | | | **6/7** | **6/7** | **4/7** | **6/7** | **5/7** | **6/7** |
| **Summary VCF ROI (reported in VCF/total indels)** | | | **1/7*** ** | **5/7*** ** | **1/7*** ** | **6/7** | **6/7** | **4/7** |

*exonic; **in our region of interest (ROI); ***note that in the VCF files of V1 only variants in designed target region are called (cf. Supplementary Table S3); NA: not applicable (here: zero reads); bold: unambiguous calls (>4 reads carrying indel reported by IGV); italic: in designed target region; unmarked: in region flanking to designed target region (within ±100 bp); underlined: outside the flanking region; [1]variant called in VCF; [2]variant not called in VCF.

**Table S21.** Genomic regions (182 exons) with copy numbers known from array CGH (highlighted in Supplementary Figure S26).

| Chr | Start | End | Length [bp] | Exons | 44 | 280 | 326* | 2905 | 7344 | 7739 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 206'315'856 | 206'407'396 | 91'541 | 9 | wt | wt | wt | het del | wt | wt |
| 4 | 69'373'934 | 69'491'021 | 117'088 | 6 | het del | het del | wt | wt | wt | het del |
| 8 | 15'950'680 | 16'023'855 | 73'176 | 7 | het del | wt | wt | wt | wt | wt |
| 8 | 39'232'101 | 39'387'546 | 155'446 | 30 | wt | het del | wt | het del | het del | hom del |
| 10 | 47'585'204 | 47'703'746 | 118'543 | 17 | dup | wt | wt | wt | wt | wt |
| 11 | 55'364'154 | 55'432'019 | 67'866 | 3 | wt | wt | wt | wt | het del | wt |
| 11 | 98'964'932 | 100'557'616 | 1'592'685 | 25 | wt | wt | wt | wt | het del | wt |
| 14 | 73'994'506 | 74'025'405 | 30'900 | 3 | wt | hom del | wt | het del | hom del | het del |
| 15 | 48'910'861 | 48'940'773 | 29'913 | 2 | het del | wt | wt | wt | wt | wt |
| 19 | 6'889'723 | 7'017'490 | 127'768 | 39 | wt | wt | wt | wt | wt | dup |
| 22 | 25'630'803 | 25'911'781 | 280'979 | 16 | het del | wt | wt | wt | wt | wt |
| 22 | 42'897'410 | 42'955'581 | 58'172 | 25 | wt | dup | wt | wt | wt | dup |

* This sample was used as control for the relative base count quantification in WES data; wt, wild-type; het del, heterozygous deletion; hom del, homozygous deletion; dup, duplication.

**Table S22.** Total raw read counts reported by the four vendors (V1-V4).

| Sample | Agilent | | NimbleGen | | Illumina | |
|---|---|---|---|---|---|---|
| | V1 | V2 | V1 | V3 | V1 | V4 |
| 44 | 154'666'636 | 188'901'528 | 223'738'732 | 164'277'220 | 106'883'310 | 68'141'762 |
| 280 | 219'518'670 | 173'953'326 | 159'813'888 | 199'864'124 | 115'773'968 | 57'109'004 |
| 326 | 137'556'636 | 173'729'966 | 154'276'308 | 142'064'946 | 95'556'070 | 71'901'946 |
| 2905 | 130'559'938 | 144'762'290 | 178'017'756 | 149'900'120 | 133'200'476 | 72'876'912 |
| 7344 | 172'323'066 | 172'157'056 | 167'946'762 | 195'763'608 | 122'058'350 | 70'667'504 |
| 7739 | 151'440'766 | 178'561'090 | 190'473'376 | 99'625'288 | 89'940'532 | 62'495'098 |
| **Mean** | **161'010'952** | **172'010'876** | **179'044'470** | **158'582'551** | **110'568'784** | **67'198'704** |

**Table S23.** Proportion of duplicates reported by the four vendors (V1-V4).

| Sample | Agilent | | NimbleGen | | Illumina | |
|---|---|---|---|---|---|---|
| | V1 | V2 | V1 | V3 | V1 | V4 |
| 44 | 26.3% | 8.0% | 10.7% | 50%* | 12.3% | 12.4% |
| 280 | 43.9% | 6.6% | 12.5% | 10%* | 12.4% | 11.1% |
| 326 | 34.7% | 8.2% | 14.0% | 50%* | 11.6% | 11.9% |
| 2905 | 24.0% | 9.5% | 8.8% | 30%* | 13.2% | 12.4% |
| 7344 | 20.2% | 8.2% | 8.7% | 30%* | 12.2% | 11.9% |
| 7739 | 19.4% | 10.0% | 9.5% | 50%* | 11.7% | 11.9% |
| **Mean** | **28.1%** | **8.4%** | **10.7%** | **37%*** | **12.2%** | **11.9%** |

*Number is estimated from provided graphs.

**Table S24.** Total mapped and deduplicated read counts (% of raw reads).

| Sample | Agilent | | NimbleGen | | Illumina | |
|---|---|---|---|---|---|---|
| | V1 | V2 | V1 | V3* | V1 | V4* |
| 44 | 110'584'430 (71.5%) | 172'118'729 (91.1%) | 186'538'676 (83.4%) | 72'899'287 (44.4%) | 88'290'788 (85.5%) | 53'245'236 (78.1%) |
| 280 | 119'682'978 (54.5%) | 160'900'330 (92.5%) | 129'934'094 (81.3%) | 158'324'581 (79.2%) | 95'545'396 (85.1%) | 45'562'196 (79.8%) |
| 326 | 87'403'593 (63.5%) | 157'585'695 (90.7%) | 122'855'749 (79.6%) | 61'382'269 (43.2%) | 79'962'090 (87.1%) | 56'364'527 (78.4%) |
| 2905 | 96'387'713 (73.8%) | 129'503'393 (89.5%) | 151'654'787 (85.2%) | 90'817'153 (60.6%) | 109'038'759 (84.8%) | 57'138'610 (78.4%) |
| 7344 | 133'734'129 (77.6%) | 156'320'331 (90.8%) | 143'047'403 (85.2%) | 121'277'235 (62.0%) | 101'342'654 (85.7%) | 55'996'771 (79.2%) |
| 7739 | 118'295'260 (78.1%) | 156'695'689 (87.8%) | 159'150'546 (83.6%) | 41'985'030 (42.1%) | 72'695'510 (87.0%) | 47'498'550 (76.0%) |
| **Mean** | **111'014'684 (69.9%)** | **155'520'695 (90.4%)** | **148'863'543 (83.0%)** | **91'114'259 (55.2%)** | **91'145'866 (85.7%)** | **52'634'315 (78.3%)** |

*Only unique reads (cf. Supplementary Table S3).

**Table S25.** Enrichment and detection of non-reference (alternative) alleles in VCF files provided by vendors (V1-V4). Fraction (%) of non-reference (alternative) alleles for shared sequence variants targeted by each platform and located within RefSeq coding exons completely (100%) covered with ≥20 reads by all six platform-vendor combinations. Analysis was performed using filtered and recalibrated (V1), filtered only (V2) or unfiltered VCF files (V3 andV4) provided by vendor V1 using the same data analysis workflow for all three platforms ensuring best comparability and vendors V2-V4 with different data analysis settings as specified in Supplementary Table S3.

| **SNVs and indels** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Total** |
|---|---|---|---|---|---|---|---|
| Number of variants | 2'631 | 1'856 | 2'079 | 2'837 | 2'915 | 1'421 | **13'739** |
| Number of heterozygous variants | 1'707 | 1'196 | 1'294 | 1'812 | 1'772 | 957 | **8'738** |
| Agilent_V1 alternative allele [%] | 47.98±0.31 | 47.89±0.33 | 48.08±0.38 | 47.96±0.32 | 48.10±0.28 | 47.97±0.38 | **47.40±0.14** |
| Agilent_V2 alternative allele [%]* | 48.41±0.25 | 48.28±0.31 | 48.23±0.29 | 48.34±0.29 | 48.36±0.26 | 48.18±0.35 | **47.60±0.12** |
| NimbleGen_V1 alternative allele [%] | 46.12±0.32 | 46.31±0.40 | 45.90±0.41 | 46.55±0.31 | 46.66±0.33 | 46.57±0.42 | **47.64±0.15** |
| NimbleGen_V3 alternative allele [%]* | 48.07±0.39 | 48.01±0.33 | 48.11±0.47 | 47.90±0.34 | 48.01±0.30 | 47.96±0.65 | **47.77±0.16** |
| Illumina_V1 alternative allele [%] | 47.90±0.37 | 47.66±0.46 | 48.59±0.45 | 47.92±0.35 | 48.00±0.36 | 48.54±0.54 | **46.86±0.17** |
| Illumina_V4 alternative allele [%]* | 46.28±0.44 | 45.56±0.60 | 45.86±0.51 | 45.99±0.42 | 46.32±0.41 | 45.80±0.58 | **46.32±0.19** |
| **SNVs** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Total** |
| Number of variants | 2'606 | 1'836 | 2'058 | 2'801 | 2'876 | 1'402 | **13'579** |
| Number of heterozygous variants | 1'697 | 1'188 | 1'289 | 1'801 | 1'761 | 951 | **8'687** |
| Agilent_V1 alternative allele [%] | 47.99±0.31 | 47.92±0.33 | 48.10±0.38 | 47.97±0.32 | 48.11±0.28 | 48.00±0.38 | **47.42±0.14** |
| Agilent_V2 alternative allele [%]* | 48.43±0.25 | 48.30±0.31 | 48.27±0.28 | 48.35±0.29 | 48.38±0.26 | 48.18±0.35 | **47.62±0.12** |
| NimbleGen_V1 alternative allele [%] | 46.16±0.32 | 46.38±0.40 | 45.96±0.41 | 46.56±0.31 | 46.73±0.33 | 46.56±0.42 | **47.68±0.15** |
| NimbleGen_V3 alternative allele [%]* | 48.07±0.39 | 48.02±0.33 | 48.12±0.47 | 47.91±0.34 | 48.02±0.30 | 47.97±0.66 | **47.78±0.16** |
| Illumina_V1 alternative allele [%] | 47.93±0.37 | 47.73±0.47 | 48.65±0.45 | 47.95±0.35 | 48.02±0.36 | 48.56±0.54 | **46.94±0.17** |
| Illumina_V4 alternative allele [%]* | 46.29±0.44 | 45.60±0.60 | 45.88±0.51 | 45.99±0.42 | 46.34±0.41 | 45.77±0.58 | **46.37±0.20** |
| **Indels** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Total** |
| Number of variants | 25 | 20 | 21 | 36 | 39 | 19 | **160** |
| Number of heterozygous variants | 10 | 8 | 5 | 11 | 11 | 6 | **51** |
| Agilent_V1 alternative allele [%] | 47.13±4.66 | 42.99±4.92 | 41.25±8.58 | 46.95±4.68 | 47.05±4.06 | 45.45±6.26 | **47.11±1.81** |
| Agilent_V2 alternative allele [%]* | 44.85±3.83 | 43.95±4.55 | 36.27±6.48 | 46.62±4.19 | 45.60±3.78 | 46.99±5.83 | **47.20±1.58** |
| NimbleGen_V1 alternative allele [%] | 40.39±4.77 | 36.63±5.88 | 32.78±9.27 | 44.21±4.53 | 35.69±4.69 | 48.87±6.99 | **47.00±1.95** |
| NimbleGen_V3 alternative allele [%]* | 48.10±5.85 | 47.70±4.83 | 44.34±10.69 | 45.20±5.02 | 47.19±4.34 | 45.69±10.82 | **47.55±2.19** |
| Illumina_V1 alternative allele [%] | 43.13±5.51 | 38.04±6.85 | 34.65±10.14 | 43.15±5.10 | 43.57±5.17 | 44.39±8.86 | **45.55±2.22** |
| Illumina_V4 alternative allele [%]* | 45.10±6.64 | 39.41±8.81 | 42.18±11.51 | 44.31±6.05 | 43.94±5.86 | 50.17±9.63 | **45.42±2.61** |

Indicated ranges for mean values (±) represent 95% confidence intervals; *since each vendor applied a different data analysis workflow (cf. Supplementary Table S3), the restriction to shared sequence variants targeted by the design of each platform and located within RefSeq coding exons completely covered at 20× by all six platform-vendor combinations should largely exclude possible false-positive allele calls.

**Table S26.** Enrichment and detection of non-reference (alternative) alleles in gVCF files generated by the same in-house bioinformatics pipeline. Fraction (%) of alternative alleles for shared sequence variants within the platforms' designed target region and 50-bp flanking sequences achieving ≥20 reads and >30 quality scores by all six platform-vendor combinations.

| SNVs and indels | 44 | 280 | 326 | 2905 | 7344 | 7739 | Total |
|---|---|---|---|---|---|---|---|
| Number of variants | 27'818 | 21'518 | 23'758 | 28'188 | 28'347 | 21'565 | **151'194** |
| Number of heterozygous variants | 18'152 | 13'971 | 15'142 | 17'876 | 18'163 | 14'499 | **97'803** |
| Agilent_V1 alternative allele [%] | 47.33±0.12 | 47.61±0.13 | 47.45±0.14 | 47.41±0.12 | 47.40±0.11 | 47.22±0.13 | **47.40±0.05** |
| Agilent_V2 alternative allele [%] | 47.58±0.11 | 47.69±0.12 | 47.64±0.12 | 47.58±0.12 | 47.56±0.11 | 47.55±0.13 | **47.60±0.05** |
| NimbleGen_V1 alternative allele [%] | 47.83±0.12 | 47.70±0.14 | 47.50±0.14 | 47.69±0.13 | 47.55±0.13 | 47.56±0.14 | **47.64±0.05** |
| NimbleGen_V3 alternative allele [%] | 47.88±0.13 | 47.38±0.12 | 47.94±0.15 | 47.73±0.12 | 47.76±0.11 | 47.88±0.18 | **47.77±0.06** |
| Illumina_V1 alternative allele [%] | 46.86±0.15 | 46.93±0.16 | 46.56±0.17 | 46.97±0.14 | 47.12±0.14 | 46.64±0.18 | **46.86±0.06** |
| Illumina_V4 alternative allele [%] | 46.53±0.16 | 46.10±0.19 | 46.04±0.18 | 46.40±0.16 | 46.50±0.16 | 46.21±0.18 | **46.32±0.07** |
| **SNVs** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Total** |
| Number of variants | 26'221 | 20'285 | 22'422 | 26'470 | 26'654 | 20'312 | **142'364** |
| Number of heterozygous variants | 17'138 | 13'163 | 14'289 | 16'793 | 17'082 | 13'693 | **92'158** |
| Agilent_V1 alternative allele [%] | 47.33±0.12 | 47.63±0.14 | 47.46±0.14 | 47.43±0.13 | 47.42±0.12 | 47.26±0.14 | **47.42±0.05** |
| Agilent_V2 alternative allele [%] | 47.60±0.11 | 47.72±0.13 | 47.65±0.12 | 47.62±0.12 | 47.59±0.11 | 47.58±0.13 | **47.62±0.05** |
| NimbleGen_V1 alternative allele [%] | 47.86±0.13 | 47.73±0.15 | 47.51±0.14 | 47.74±0.13 | 47.60±0.13 | 47.63±0.14 | **47.68±0.06** |
| NimbleGen_V3 alternative allele [%] | 47.91±0.14 | 47.40±0.12 | 47.96±0.16 | 47.74±0.13 | 47.78±0.11 | 47.86±0.18 | **47.78±0.06** |
| Illumina_V1 alternative allele [%] | 46.92±0.15 | 47.01±0.17 | 46.66±0.17 | 47.05±0.15 | 47.18±0.15 | 46.72±0.18 | **46.94±0.07** |
| Illumina_V4 alternative allele [%] | 46.58±0.16 | 46.17±0.20 | 46.08±0.18 | 46.47±0.16 | 46.53±0.16 | 46.28±0.19 | **46.37±0.07** |
| **Indels** | **44** | **280** | **326** | **2905** | **7344** | **7739** | **Total** |
| Number of variants | 1'597 | 1'233 | 1'336 | 1'718 | 1'693 | 1'253 | **8'830** |
| Number of heterozygous variants | 1'014 | 808 | 853 | 1'083 | 1'081 | 806 | **5'645** |
| Agilent_V1 alternative allele [%] | 47.41±0.55 | 47.28±0.57 | 47.23±0.61 | 47.11±0.53 | 47.04±0.47 | 46.53±0.62 | **47.11±0.23** |
| Agilent_V2 alternative allele [%] | 47.34±0.50 | 47.30±0.54 | 47.39±0.53 | 47.07±0.50 | 47.09±0.45 | 47.07±0.59 | **47.20±0.21** |
| NimbleGen_V1 alternative allele [%] | 47.86±0.55 | 47.73±0.60 | 47.51±0.59 | 47.74±0.51 | 47.60±0.51 | 47.63±0.62 | **47.00±0.23** |
| NimbleGen_V3 alternative allele [%] | 47.44±0.58 | 47.11±0.51 | 47.57±0.65 | 47.55±0.51 | 47.41±0.47 | 48.31±0.75 | **47.55±0.23** |
| Illumina_V1 alternative allele [%] | 45.71±0.69 | 45.62±0.76 | 44.86±0.78 | 45.66±0.62 | 46.04±0.60 | 45.21±0.82 | **45.55±0.29** |
| Illumina_V4 alternative allele [%] | 45.73±0.71 | 44.85±0.87 | 45.37±0.81 | 45.25±0.70 | 46.06±0.66 | 45.06±0.85 | **45.42±0.31** |

Variant positions with more than one different non-reference allele (non-biallelic) as well as variant calls with alternative allele percentages outside 10-90% were excluded; indicated ranges for the mean values (±) represent 95% confidence intervals.

**Table S27.** Array CGH data for three DNA samples (44, 7344, and 7739). Common coding SNVs of NimbleGen CGH/LOH array within designed target region and exons completely covered at ≥20 reads by WES performed by all platforms and vendors compared to WES variant calls (unfiltered VCF files provided by vendors, cf. Supplementary Table S3).

| | **44** | **7344** | **7739** |
|---|---|---|---|
| Total shared covered coding exons | 65'135 | 68'992 | 39'704 |
| Total SNVs from array CGH in covered exons | 93 | 101 | 53 |
| Heterozygous SNPs | 44* | 29* | 20* |
| Homozygous SNPs | 10* | 23* | 6* |
| Wild-type SNPs | 27** | 41** | 13** |
| No/false array results | 12 | 8 | 14 |

*Correctly called regardless of WES platform and vendor; **correctly recognised as wild-type regardless of WES platform and vendor.

# Appendix 3 Selected Candidate Genes for AD

These 226 AD candidate genes were selected according to literature, meeting abstracts, and mouse models. These genes were in more detail analysed in aCGH and were included in our AD candidate gene panel for NGS (status June 2014).

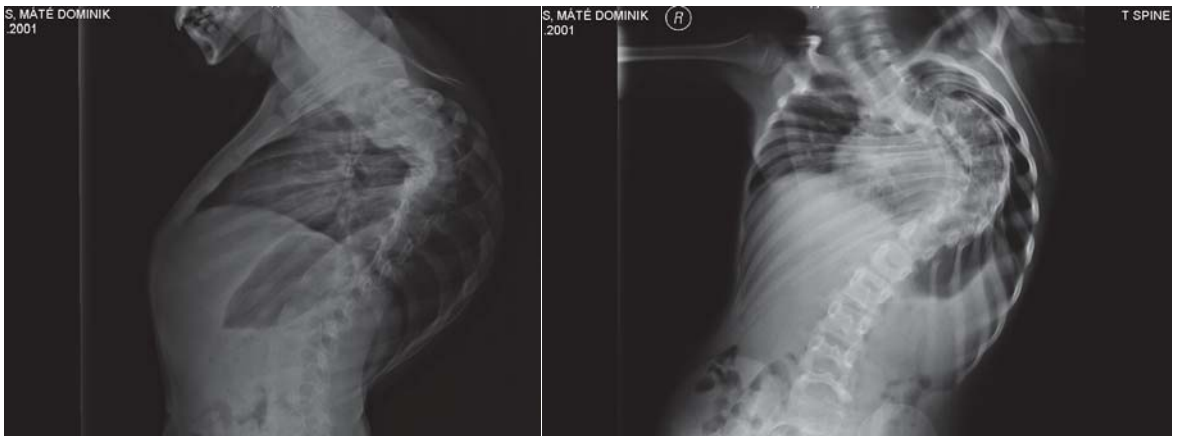
*ABCC6, ACTA2, ACVRL1, ADAM10, ADAM12, ADAM15, ADAM17, ADAMTS10, ADAMTS17, ADAMTS2, ADAMTSL2, ADAMTSL4, AGTR2, AKT2, ANXA2, APC, APOE, ATP7A, B3GAT1, BGN, BHMT, BHMT2, BMP1, BMP10, BMP15, BMP2, BMP3, BMP4, BMP5, BMP6, BMP7, BMP8A, BMP8B, BRAF, CAPN2, CBS, CCM2, CDC25A, CDKN2A, CDKN2B, CFC1, CHD7, CHST14, CIB3, COL11A1, COL11A2, COL16A1, COL18A1, COL1A1, COL1A2, COL2A1, COL3A1, COL4A1, COL4A2, COL4A5, COL5A1, COL5A2, COL9A1, COL9A2, COL9A3, CSRP2, CTGF, CYBB, DCHS1, DCN, DSE, EBP, EFEMP1, EFEMP2, EFNB1, ELN, EMILIN1, ENG, ETS1, FBLN1, FBLN5, FBN1, FBN2, FGF10, FGF8, FGFR2, FKBP14, FLCN, FLNA, FMOD, FN1, GAA, GATA6, GDF1, GDF11, GDF2, GDF5, GDF6, GDF7, GJA1, GLA, GPR4, HAPLN1, HAS1, HMGA2, HRAS, HSPB1, HSPG2, IL1B, IL7, ITGA1, ITGAV, ITGB1, ITGB3, JAG1, JAG2, KCNJ2, KLF15, KLF2, KLK1, KRAS, KRIT1, LEFTY2, LEMD3, LEPRE1, LEPREL1, LOX, LOXL1, LRP1, LRRC7, LTBP2, LTBP3, LTBP4, LUM, MAGI2, MAP2K1, MAP2K2, MAPK1, MAPK3, MED12, MFAP1, MGP, MLL2, MMP1, MMP10, MMP12, MMP14, MMP2, MMP3, MMP8, MMP9, MYH11, MYLK, NCF1, NF1, NKX2-5, NKX2-6, NMRK2, NOS1, NOS3, NOTCH1, NOTCH2, NOX1, NPHP3, NR3C1, NRAS, PCOLCE, PDCD10, PIK3C3, PKD1, PKD2, PLAU, PLOD1, PLOD2, PLOD3, PPIB, PRKG1, PTGS1, PTGS2, PTPN11, RAF1, RAG1, RAG2, RAI1, RCN2, REN, RTN1, RTN4, S100A12, SERPINE1, SHOC2, SKI, SLC2A10, SLC39A13, SLC7A1, SMAD1, SMAD2, SMAD3, SMAD4, SMAD5, SMAD6, SMAD7, SMAD9, SOS1, SOX4, SPOCK2, TAGLN, TAZ, TBRG1, TBX1, TGFB1, TGFB2, TGFB3, TGFBR1, TGFBR2, TGFBR3, TGM2, THBS1, THBS2, THSD4, TIMP1, TIMP2, TIMP3, TIMP4, TNXB, TRBV4-1, TRBV4-2, TRBV4-3, TSC2, VCAM1, VCAN*

# Appendix 4 Additional aCGH Project

## Rare Disease Case – Dominik (11-year-old boy)

### Dominik's Main Clinical Symptoms

- Muscle hypotonia, ataxia
- Therapy-resistant epilepsy, tonic seizures
- Developmental delay
- Mental retardation
- **Severe kyphoscoliosis**
- Right ventricular hypertrophy
- Gluten hypersensitivity




Dominik has needed an immediate surgical intervention of his severe kyphoscoliosis (KISPI and Schulthess Klinik Zurich)

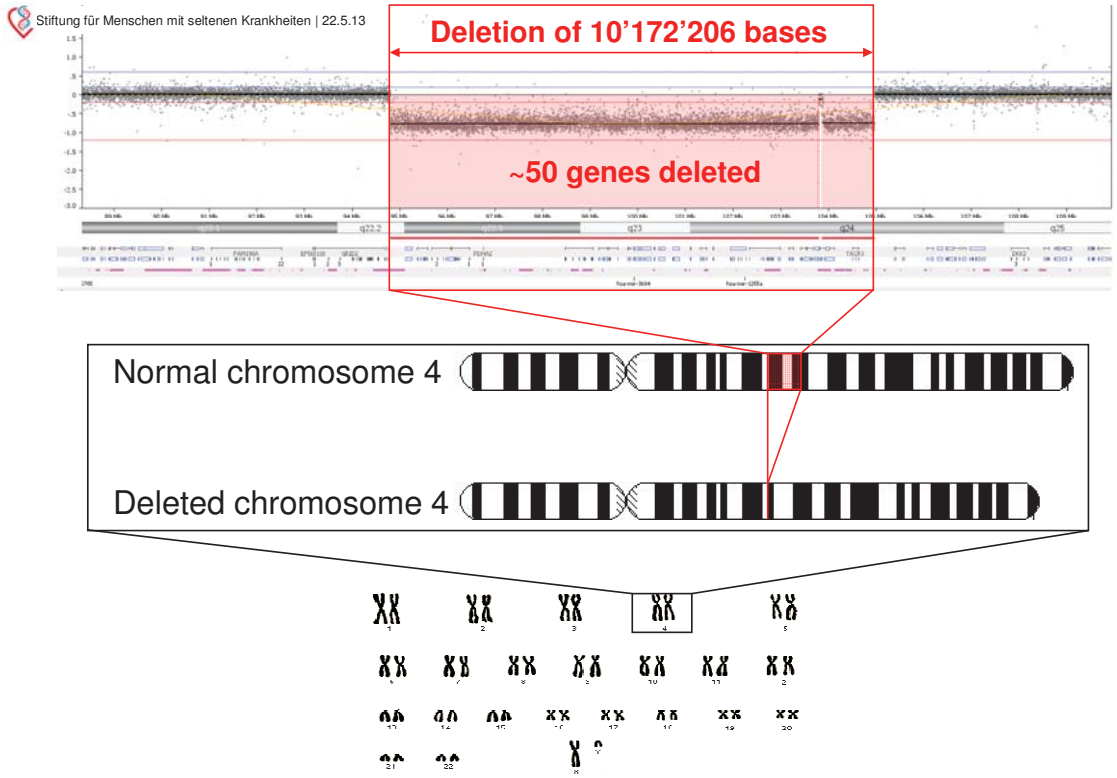
Stiftung für Menschen mit seltenen Krankheiten | 22.5.13

### Dominik's Medical History

Stiftung für Menschen mit seltenen Krankheiten | 22.5.13

He was born with a relatively big umbilical hernia corrected with surgery on day 2.
6. month - severe hypotonia and scoliosis developed, he was put in a brace
12. month - DSGM therapy to treat hypotonia leading to an effective recovery
21. month - he could stand up
29. month - he could walk alone safely
31. month - EEG: normal background activities for his age, paroxysmal signals could not be detected, epilepsy could be excluded
38. month - end of DSGM therapy
49. month - epileptic seizures (myoclonus) developed, EEG: severe cortical functional disturbances, multifocal diffuse interictal paroxysmal activities
49.-61. months - trying various antiepileptic drugs without any success, his walk became less secure, more frequent seizures and absence periods, therapy resistant reflex epilepsy developed, stereotype behaviors started when playing with toys
61.-64. months - ketogenic diet, he became very weak with autistic signs and lost mobility, soon after his diet was stopped he regained some of his former physical skills
64.-68. months - all available biochemical and genetic testing were done, all negative
76.-84. months - antiepileptic drugs were gradually stopped, introduced various herbal extracts and intensive physiotherapies, reflex epilepsy could be controlled and he regained his motility skills
9. year - gluten hypersensitivity, gluten-free diet, dietary supplements, natural antiepileptics

**11. year (March, 2013) – genetic testing revealed the deletion of ~50 genes of chr. 4**
(performed by the *Stiftung für Menschen mit seltenen Krankheiten*)




**Dominik needs help** for the identification which of the ~50 deleted genes is/are responsible for his main symptoms, requiring further genetic analyses and targeted clinical investigations

## Rare Disease Case – Dominik (11-year-old boy)

### Dominik's main clinical symptoms

- Muscle hypotonia, ataxia
- Therapy-resistant epilepsy, tonic seizures
- Developmental delay
- Mental retardation
- Severe kyphoscoliosis
- Right ventricular hypertrophy
- Gluten hypersensitivity

### Disease cause

Non-inherited (*de novo*) 10-Mb deletion of the long arm of chromosome 4 affecting ~50 genes (March, 2013, identified by the *Stiftung für Menschen mit seltenen Krankheiten*)

### Next step

To develop a targeted therapy based on the molecular basis of his clinical symptoms, replacing/extending his **recent therapies (gluten-free diet, alternative antiepileptic treatments, and the so-called DSGM therapy [complex motion rehabilitation process])**

# Appendix 5 Application to Perform Animal Experiments

Tierversuche Form. **A** Nr. .....
(von der Bewilligungsstelle auszufüllen)

## Gesuch für Tierversuche

1 Adresse Gesuchsteller/in (Kontaktperson, Institut, Firma)
**PD Dr. Gabor Matyas**
Zentrum für Kardiovaskuläre Genetik & Gendiagnostik
Wagistrasse 25
8952 Schlieren
E-mail; Tel-Nr. matyas@genetikzentrum.ch; 043 433 86 86

2 Adresse der kantonalen Bewilligungsstelle
Veterinäramt des Kantons Zürich
Obstgartenstrasse 21
8090 Zürich

3 **TITEL DES GESUCHS / PROJEKTS**
Überprüfung der Wirkung von verschiedenen Substanzen in Mausmodellen für Ehlers-Danlos Syndrom vaskulärer Typ (EDS IV), Kurztitel: Therapieansätze EDS IV Maus

31 Fachgebiet(e) bzw. Anwendungsbereich(e):
Humangenetik, Bindegewebe

32 ☒ *[N]* neues Gesuch
☐ *[F]* Fortsetzungsgesuch (zu Nr. )
☐ *[E]* Ergänzungsgesuch* (zu Nr. ) → Änderung ☐ methodischer Art ☐ Tierzahl /Tierart ☐ Gültigkeit /Verlängerungsdauer ☐ anderes (**betreffende Ziffern im Gesuch nennen**)

*Bei Ergänzungsgesuchen bitte nur Änderungen notieren oder Änderungen markieren

33

| **TIERART (Stamm)** | **Gesamtzahl** pro Gesuch | **Herkunft*** (a-c) |
|---|---|---|
| Col3 alpha 1 delta Wildtyp | 370 (s. 54.3) | b |
| Col3 alpha 1 delta Heterozygot | 810 (s. 54.3) | b |
| C.129S4(B6)-Col3a1tm1Jae/J Wildtyp | 110 (s. 54.3) | b |
| C.129S4(B6)-Col3a1tm1Jae/J Heterozygot | 170 (s. 54.3) | b |

*Herkunft: (a): aus früherem Versuch, welche:
(b): anerkannte Versuchstierzucht oder -handlung (Art. 59b, TSchV);
(c): andere Herkunft, welche: MRC Harwell

Namen und Adressen der Lieferanten:
MRC Harwell, Harwell Science and Innovation Campus, Oxfordshire, OX11 0RD, UK (Col3 alpha 1 delta)
Prof. Dr. Thierry Hennet, Physiologisches Institut, Winterthurerstr. 190, 8057 Zürich (C.129S4(B6)-Col3a1tm1Jae/J)

34 **Adresse Tierhaltungsort:**
BioSupport AG, Wagistrasse 10, 8952 Schlieren

**Verwendung gentechnisch veränderter Tiere*** : ja ☒ nein ☐
*Datenblätter zur Erfassung und Charakterisierung gentechnisch veränderter Tiere sind beizufügen.

35 **Maximal erwarteter Schweregrad**: 3 (Details: siehe unter Ziff. 56.4)

36 **Dauer des gesamten Vorhabens:** 3 Jahre
**Datum des vorgesehenen Beginns:** 1. Juli/November 2012

37 **Liste der Personen, die Massnahmen und Eingriffe im Rahmen des Versuchs durchführen oder leiten:**
Hauptversuchsleiter: Dr. Steffen M. Zeisberger
Stellvertretender Tierversuchsleiter: Prof. Dr. Manfred Kopf
Versuchsdurchführende Person: MSc (Humanbiologie) Janine Meienberg (PhD-Studentin)

Verteiler: Original des Gesuchs und mind. 3 Kopien an die Bewilligungsbehörde

**Gesuch** (Form. A) **- 2 -** Nr. .....
(von der Bewilligungsstelle auszufüllen)

**Kurztitel:** Überprüfung der Wirkung von verschiedenen Substanzen in Mausmodellen für Ehlers-Danlos Syndrom vaskulärer Typ (EDS IV), Kurztitel: Therapieansätze EDS IV Maus

Personen, die Versuche leiten oder durchführen, müssen die Aus- und Weiterbildungsanforderungen gemäss der „Verordnung über die Aus- und Weiterbildungen des Fachpersonals für Tierversuche" vom 12. Oktober 1998 erfüllen.
Der entsprechende Anhang „beteiligte Personen und Nachweis der Aus- und Weiterbildung" ist auszufüllen und die Aus- und Weiterbildungsnachweise sind als Kopien beizulegen.

38 Der/die unterzeichnete **wissenschaftliche Instituts- oder Laborleiter/in** (Art. 61a, Abs. 1, TSchV) bestätigt, dass den unter **Ziff. 37** genannten Personen die auf Tierversuche anwendbaren Vorschriften des TSchG und der TSchV bekannt sind und dass sie die Aus- und Weiterbildungsanforderungen erfüllen (Art. 59d, TSchV).

| Ort und Datum | Name und Unterschrift |
|---|---|
| 15.06.2012 | PD Dr. Gabor Matyas |

39 **Versuchsleiter/in**: falls mehrere Personen diese Funktion ausüben, ist deren Verantwortungsbereich gemäss Ziff. 37 festzulegen. **Unterschrift des Hauptversuchsleiters/der Hauptversuchsleiterin**: (Art. 59d, TSchV):

| Ort und Datum | Name und Unterschrift |
|---|---|
| 15.06.2012 | Dr. Steffen M. Zeisberger |

4 **ANGABEN ZUR FRAGESTELLUNG ODER ZIELSETZUNG** (für die Statistik Art. 64b, TSchV); Ziff. 41 - 43 **ausschliesslich je einmal** ankreuzen und ggf. bei den Detailfragen eine weitere Marke sowie Ergänzungen anbringen.

41 Das Vorhaben steht in Zusammenhang mit

- ☒ biologischen (einschliesslich medizinischen) Untersuchungen im Bereich der Grundlagenforschung
- ☐ Entdeckung, Entwicklung und Qualitätskontrolle (exkl. Unbedenklichkeitsprüfung) von Produkten oder Geräten in der Human- oder Veterinärmedizin
- ☐ Krankheitsdiagnostik
- ☐ Bildung und Ausbildung
- ☐ dem Schutz von Mensch, Tier und Umwelt durch toxikologische oder sonstige Unbedenklichkeitsprüfungen

  ... für Stoffe, die überwiegend
  - ☐ als Arzneimittel (einschliesslich medizinischer Geräte)
  - ☐ in der Landwirtschaft
  - ☐ in der Industrie
  - ☐ in Privathaushalten
  - ☐ als Kosmetik- oder Toilettenartikel
  - ☐ als Lebensmittel-Zusatzstoffe verwendet werden oder zu einer solchen Verwendung bestimmt sind oder
  - ☐ der Abklärung von möglichen oder tatsächlichen Gefahren von Kontaminanten in der allgemeinen Umwelt dienen, oder
  - ☐ andere Verwendung. Welche:

- ☐ anderer Zusammenhang. Welcher:

42 Das Vorhaben steht in Zusammenhang mit

- ☒ Krankheiten beim Menschen
  - ☐ Krebs (mit Ausnahme der Kanzerogenitätsprüfungen)
  - ☐ Herz-Kreislauf-Erkrankungen
  - ☐ Nerven- und Geistesstörungen
  - ☒ sonstige Krankheiten beim Menschen. Welche: vererbte Bindegewebskrankheiten
- ☐ Krankheiten beim Tier. Welche:
- ☐ Kein Zusammenhang mit Krankheiten bei Mensch und Tier.

43 Das Vorhaben steht in Zusammenhang mit gesetzlich vorgesehenen Verfahren (Registrierungs- und Zulassungsvorschriften):

- ☐ nur für die Schweiz
- ☐ nur für andere Länder. Welche/s:
- ☐ Beides. Welche Länder:
  Angabe der Richtlinie/n oder Prüfvorschrift/en:

- ☒ Kein Zusammenhang mit gesetzlich vorgesehenen Verfahren.

Verteiler: Original des Gesuchs und mind. 3 Kopien an die Bewilligungsbehörde

**Gesuch** (Form. A) **- 3 -** Nr. .....
(von der Bewilligungsstelle auszufüllen)

**Kurztitel:** Überprüfung der Wirkung von verschiedenen Substanzen in Mausmodellen für Ehlers-Danlos Syndrom vaskulärer Typ (EDS IV), Kurztitel: Therapieansätze EDS IV Maus

44.1 Beschreibung des Versuchsziels (beispielsweise Zusammenfassung des NF-Gesuchs, **maximal eine Seite**):

Ehlers-Danlos Syndrom vaskulärer Typ (EDS IV) ist eine autosomal dominant vererbte Bindegewebskrankheit mit einer Prävalenz von 1-2 in 100'000 Individuen (Germain 2007, Orphanet J Rare Dis 2:32). Die Symptome umfassen dünne, durchscheinende Haut, erhöhte Anfälligkeit für blaue Flecken, typische Gesichtszüge und fragile Wände von Hohlorganen und grösseren Arterien, was zu einem erhöhten Risiko für Rupturen führt (Beighton et al. 1997, Am J Med Genet 77:31–37). So ist die schwerwiegendste Komplikation dieser Krankheit das erhöhte Risiko für Dissektionen und daraus resultierende Rupturen der Aorta und grossen Arterien und damit meist plötzliche Todesfälle. EDS IV wird durch Mutationen im *COL3A1*-Gen verursacht, welches für die alpha1-Kette von Kollagen Typ III, einem fibrillären Kollagen, kodiert (z.B. Pope et al 1975, Proc Natl Acad Sci USA 72:1314–1316). Dieses wird in den Wänden der Hohlorgane, wie grossen Blutgefässen, Blase, Darm und Uterus exprimiert. In den meisten Fällen werden missense Mutationen gefunden, welche zum Austausch eines Glycins führen. Weitere missense Mutationen, Splice Site Mutationen oder kleine Deletionen oder Insertionen konnten ebenfalls nachgewiesen werden. Nur in ganz seltenen Fällen wurden nonsense Mutationen beschrieben, welche zu funktioneller Haploinsuffizienz führen, und wir konnten erst letztes Jahr (2010) den ersten Fall von echter Haploinsuffizienz beschreiben (Meienberg et al. 2010, Eur J Hum Genet 18:1315-1321). Haploinsuffizienz bedeutet, dass es nur von einem der zwei Allele ein Proteinprodukt gibt und somit die Gesamtmenge an Protein reduziert ist. Bis jetzt gibt es keine zielgerichtete Therapie für Patienten mit EDS IV sondern nur Krankheitsmanagement und Symptombehandlung. Bei einer echten Haploinsuffizienz ist der Vorteil, dass es „nur" zu wenig Protein hat, aber kein defektes Protein. Dies macht die gezielte Therapie einfacher, da „nur" die Proteinmenge erhöht werden muss. Für EDS IV gibt es ausserdem nur Mausmodelle, welche ein Knockout-Allel von *Col3a1* haben und somit im heterozygoten Zustand echte Haploinsuffizienz für *Col3a1*. Das heisst, dass zurzeit Tierversuche zu EDS IV nur mit diesem Gendefekt gemacht werden können. Das primäre Ziel dieses Projektes ist es nun anhand dieser Mausmodelle, eine Therapie für die von uns beschriebene echte Haploinsuffizienz für *COL3A1* zu finden, welche die Gesamtmenge an Kollagen Typ III erhöht und somit zu besserer Stabilität der Wände der Hohlorgane führt und das Risiko für Rupturen senkt. In einem zweiten Schritt wird dann geschaut, ob und wie die gefundenen Substanzen und gewonnen Kenntnisse auf die weiteren Fälle von EDS IV und auch auf verwandte Aortenkrankheiten angewendet werden können, mit dem Ziel, möglichst vielen Menschen helfen zu können. In diesem Tierversuch werden wir ein erst kürzlich beschriebenes, spontan entstandenes Mausmodell, welches eine echte Haploinsuffizienz von *Col3a1* hat und bei heterozygoten Tieren wie bei den Menschen zu einer erhöhten Mortalitätsrate wegen Aortendissektionen führt, verwenden (Smith et al. 2011, Cardiovasc Res 90:182-190). Wir werden den Mäusen verschiedene, bereits in anderem Zusammenhang für den Menschen zugelassene Substanzen verabreichen, mit dem Ziel, eine Substanz zu finden, welche die mechanische Stabilität der Aortenwand erhöht und dadurch die Mortalität infolge von Aortendissektion reduziert. Die gefundene Substanz dürfte einen Therapieansatz für die von uns beschriebene Familie mit echter Haploinsuffizienz von *COL3A1*, aber auch für Patienten mit anderen Mutationen im *COL3A1* und mit anderen verwandten Krankheiten darstellen.

*Ehlers-Danlos syndrome vascular type (EDS IV) is an autosomal dominant connective tissue disorder with a prevalence of 1-2 in 100'000 individuals (Germain 2007, Orphanet J Rare Dis 2:32). It is characterized by thin translucent skin, easy bruising, and typical facial features as well as by fragile walls of hollow organs and larger arteries, which leads to an increased risk for rupture (Beighton et al. 1997, Am J Med Genet 77:31–37). Consequently, the most severe complication is the increased risk for dissections and subsequent ruptures of the aorta and large arteries, leading to sudden death. EDS IV is caused by mutations in the gene COL3A1, which encodes the alpha 1 chain of type III collagen, a fibrillar collagen (e.g. Pope et al. 1975, PNAS 72:1314–1316). It is expressed in walls of hollow organs, such as large blood vessels, bladder, bowel, and uterus. In the majority of EDS IV cases, missense mutations leading to glycine substitution can be found. Other missense or splice site mutations or small deletions/insertions can be detected as well. Only few nonsense mutations leading to functional COL3A1 haploinsufficiency have been reported and the first case of true COL3A1 haploinsufficiency has very recently been described by us (Meienberg et al. 2010, Eur J Hum Genet 18:1315-1321). Haploinsufficiency means that only one of the two alleles leads to a protein and thus the total amount of protein is reduced. So far, there is no targeted therapy for EDS IV patients available, but only disease management and treatment of symptoms. The advantage of true haploinsufficiency is that there is "only" not enough protein, but no truncated protein. This makes targeted therapy easier as "only" the amount of protein has to be increased. Furthermore, for EDS IV only two mouse models exist, which have a knockout allele of Col3a1 and thus in the heterozygous state true Col3a1 haploinsufficiency. This means that at the moment animal experiments for EDS IV can only be done with this gene defect. The primary goal of this project is to use these two mouse models in order to find a therapy for true COL3A1 haploinsufficiency by increasing the total amount of type III collagen and therefore improves the mechanical stability of the walls of hollow organs as well as reduces risk for ruptures. In a second step, we will evaluate if and how the best substance(s) and acquired knowledge can be applied to further cases of EDS IV and related aortic disorders as well, with the goal to be able to help as many people as possible. In this animal experiment, we will use a recently described, spontaneously developed mouse model, which has true haploinsufficiency of Col3a1 and leads in heterozygous animals, like in humans, to an increased mortality rate due to aortic dissection (Smith et al. 2011, Cardiovasc Res 90:182-190). We will administer different substances to mice with true Col3a1 haploinsufficiency, which are already approved for humans for other purposes, with the goal to find substance(s), which will increase the mechanical stability of the aortic wall and like this, reduce mortality due to aortic dissection. The best substance is expected to represent a therapeutic approach for the family with true COL3A1 haploinsufficiency described by us as well as for patients with different types of COL3A1 mutations or suffering from other related disorders.*

Verteiler: Original des Gesuchs und mind. 3 Kopien an die Bewilligungsbehörde

**Gesuch** (Form. A) **- 4 -** Nr. .....
(von der Bewilligungsstelle auszufüllen)

**Kurztitel:** Überprüfung der Wirkung von verschiedenen Substanzen in Mausmodellen für Ehlers-Danlos Syndrom vaskulärer Typ (EDS IV), Kurztitel: Therapieansätze EDS IV Maus

44.2 Grundsätzliche Erkenntnis, die angestrebt wird; Stand der Forschung; Darlegung, was noch nicht hinreichend bekannt ist:

EDS IV führt trotz Fortschritten in den diagnostischen und chirurgischen Möglichkeiten immer noch zu einem erheblichen Risiko von schweren Komplikationen und erhöhter Mortalität. Bis jetzt gibt es noch keine zielgerichtete Therapie. Vor kurzem wurde eine erste klinische Studie mit einem blutdrucksenkenden Mittel abgeschlossen, bei der es aber keine Untersuchung auf einen Effekt auf die Struktur der Aortenwand gibt (Ong et al. 2010, Lancet 376:1476-84). Bis vor kurzem gab es nur ein *Col3a1* Mausmodell, welches heterozygot nur einen schwach ausgeprägten Phänotypen zeigt und auch nicht zu einer erhöhten Mortalitätsrate führt (C.129S4(B6)-Col3a1tm1Jae/J; Cooper et al. 2010, Vet Pathol 47:1028). Mit diesem Modell konnte in einer kürzlich erschienenen Studie gezeigt werden, dass eine Behandlung mit Doxycyclin zu einer erhöhten Stabilität der Aortenwand führt (Briest et al. 2011, J Pharmacol Exp Ther 337:621-7). Unsere Versuche werden wir primär mit einem neuen Mausmodell für EDS IV durchführen (Col3 alpha 1 delta; Smith et al. 2011, Cardiovasc Res 90:182-190), bei welchem der kardiovaskuläre Phänotyp viel ausgeprägter ist und mehr dem der Patienten mit EDS IV entspricht. Wir wollen nun einerseits testen, ob Doxycyclin beim neuen Mausmodell eine vergleichbare Wirkung hat, wie beim alten Modell. Ausserdem wollen wir weitere Substanzen auf ihre Wirkung auf die Überlebensrate der Tiere, die Stabilität der Aorta und die Kollagenmenge in der Aortenwand untersuchen. Dies können Substanzen sein, welche bei verwandten Krankheiten das Risiko für kardiovaskuläre Komplikationen senken konnten, wie beispielsweise Losartan, welches ein Angiotensin-II-Typ-1-Rezeptor-Antagonist ist und dessen Wirksamkeit bei Marfan Syndrom (MFS) gezeigt wurde (Habashi et al. 2006, Science 312:117-21), oder Dexamethason, welches bei Fibroblasten von Patienten mit Loeys-Dietz Syndrom eine positive Wirkung zeigte (Barnett et al. 2011, Eur J Hum Genet 19:624-33). Weiter werden wir Substanzen testen, bei welchen gezeigt wurde, dass sie Matrix Metalloproteinasen (MMPs) hemmen, welche für den Abbau von Kollagen verantwortlich sind. Dies sind beispielsweise Doxycyclin, dessen Wirkung ebenfalls bei anderen Aneurysma Krankheiten wie MFS gezeigt wurde (Xiong et al. 2008, J Vasc Surg 47:166-72),

*Unpublished data*

*Despite advances in diagnostic and surgical possibilities, EDS IV still leads to a substantial risk for severe complications and increased mortality. So far, no targeted therapy is available. Recently, a first clinical trial with an antihypertensive drug was finished, in which, however, effects on structure of the aortic wall were not examined (Ong et al. 2010, Lancet 376:1476-84). Until recently, only a Col3a1 mouse model has been available, which displayed only a weakly pronounced phenotype in heterozygous state and also did not lead to an increased mortality rate (C.129S4(B6)-Col3a1tm1Jae/J; Cooper et al. 2010, Vet Pathol 47:1028). For this mouse model, a recently published study showed that treatment with doxycycline leads to an increased stability of the aortic wall (Briest et al. 2011, J Pharmacol Exp Ther 337:621-7). We will primary perform our experiments with a new, very recently introduced mouse model for EDS IV (Col3 alpha 1 delta; Smith et al. 2011, Cardiovasc Res 90:182-190), in which the cardiovascular phenotype is much more pronounced and corresponds more to that of EDS IV patients. On one hand, we intend to test whether doxycycline has a similar effect in the new mouse model compared to the old one. In addition, we also intend to examine further substances on their effect on survival rate of the animals, mechanical stability of the aorta as well as amount of collagen in the aortic wall. These could be substances, which were shown to reduce the risk for cardiovascular complications in related disorders, such as losartan, an angiotensin II type 1 receptor antagonist, which has been shown to reduce/prohibit aortic dilatation in Marfan syndrome (MFS) (Habashi et al. 2006, Science 312:117-21), or dexamethasone, which showed a positive effect on fibroblasts of patients with Loeys-Dietz syndrome (Barnett et al. 2011, Eur J Hum Genet 19:624-33). Furthermore, we will test substances, which have been shown to inhibit matrix metalloproteinases (MMPs), which are responsible for the degradation of collagens. These are for example doxycycline, whose efficiency has also been shown for the aortic disorder MFS (Xiong et al. 2008, J Vasc Surg 47:166-72),*

*Unpublished data*

Verteiler: Original des Gesuchs und mind. 3 Kopien an die Bewilligungsbehörde

**Gesuch** (Form. A) **- 5 -** Nr. .....
(von der Bewilligungsstelle auszufüllen)

**Kurztitel:** Überprüfung der Wirkung von verschiedenen Substanzen in Mausmodellen für Ehlers-Danlos Syndrom vaskulärer Typ (EDS IV), Kurztitel: Therapieansätze EDS IV Maus

5 **ANGABEN ZUR METHODE** (Beschreibungen und Anmerkung zu den Ziff. 51 - 58)

51.1 Versuchsanordnung (Übersicht über die Methode, Ablauf des Versuchsvorhabens, ggf. Name des Tiermodells)

Das verwendete Mausmodell (Col3 alpha 1 delta) zeigt echte Haploinsuffizienz für *Col3a1* (nur eine Kopie). Dies führt bei heterozygoten Tieren zu einer verminderten Stabilität der Aorta und in ~28% zu spontanen Rupturen der Aorta aufgrund von Dissektionen. Die grösste Sterblichkeit wird im Alter von 4-10 Wochen beobachtet. Das Ziel ist es nun, bei verschiedenen Substanzen zu testen, ob sie die Stabilität der Aorta erhöhen und die Mortalitätsrate senken können.

Für die Bestätigung der Resultate werden die vielversprechendsten Substanzen in einem zweiten Teil des Versuches auch noch beim zweiten, bekannteren Mausmodell für *Col3a1* (C.129S4(B6)-Col3a1tm1Jae/J) getestet. Dieses Mausmodell hat ebenfalls eine echte Haploinsuffizienz für *Col3a1*, jedoch einen milderen Phänotyp, welcher nicht zu einer erhöhten Mortalität führt. Damit soll gezeigt werden, ob die Wirkung der Substanzen unabhängig vom genetischen Background und der exakten Mutation ist. Zusätzlich werden im zweiten Teil des Versuches auch verschiedene Dosen und Zeitspannen der Medikation getestet.

**Pilotversuch**

Um festzustellen, ab welchem Alter der Mäuse ein klarer Unterschied in der mechanischen Stabilität der Aorta messbar ist, wird diese bei heterozygoten Col3 alpha 1 delta und bei wildtyp Mäusen zu verschiedenen Zeitpunkten gemessen. Dazu wird eine Methode verwendet, welche mit der von Cooper et al. 2010 (Vet Pathol 47:1028) beschriebenen Methode vergleichbar ist. Anhand dieser Resultate kann dann das Alter bei der Euthanasierung im eigentlichen Versuch (s. unten) bestimmt werden. Beim zweiten Mausmodell (Col3a1tm1Jae) wird dann, bevor es im Teil 2c zum Einsatz kommt, gemessen, ob es bei dem für die Col3 alpha 1 delta Mäuse gefundenen Alter auch schon einen signifikanten Unterschied in der mechanischen Stabilität gibt und ob die von uns gemessenen Werte mit denjenigen von Cooper et al. 2010 (Vet Pathol 47:1028) vergleichbar sind (signifikante Unterschiede im Alter von 21 Monaten).

**Versuch Teil 1**

Testen von verschiedenen Substanzen auf ihre Wirkung, d.h. auf die Mortalitätsrate, die mechanische Stabilität der Aorta und die *Col3a1*-Expression in heterozygoten Tieren im Vergleich zu unbehandelten Heterozygoten und zu WT.

- 3 Gruppen von Tieren: WT unbehandelt, Col3 alpha 1 delta heterozygot unbehandelt, Col3 alpha 1 delta heterozygot behandelt.
- Die Konzentration der Substanzen basiert auf Angaben in früheren Publikationen mit anderen Fragestellungen und wird so festgelegt, dass keine Nebenwirkungen auftreten sollten.
- Nach einem bestimmten Behandlungszeitraum (gemäss Resultaten aus dem Pilotversuch) werden die Tiere euthanasiert, die mechanische Stabilität der Aorta bestimmt und Aortenproben für Histologie und RNA-Extraktion entnommen.

**Substanzen:**

- **Metalloproteinasen-Hemmer**
  - Doxycyclin (in subantimikrobiellen Konzentrationen):
    Mit dem älteren Modell für EDS IV (Col3a1tm1Jae) konnte gezeigt werden, dass eine Behandlung mit Doxycyclin, einem Tetracyclin-Antibiotikum und Metalloproteinase-Inhibitor, zu einer erhöhten Stabilität der Aortenwand führt (Briest et al. 2011, J Pharmacol Exp Ther 337:621-7). Zudem konnte die positive Wirkung von Doxycyclin ebenfalls bei anderen Aneurysma Krankheiten wie MFS gezeigt werden (Xiong et al. 2008, J Vasc Surg 47:166-72).

*Unpublished data*

Verteiler: Original des Gesuchs und mind. 3 Kopien an die Bewilligungsbehörde

**Gesuch** (Form. A) **- 6 -** Nr. .....
(von der Bewilligungsstelle auszufüllen)

**Kurztitel:** Überprüfung der Wirkung von verschiedenen Substanzen in Mausmodellen für Ehlers-Danlos Syndrom vaskulärer Typ (EDS IV), Kurztitel: Therapieansätze EDS IV Maus

- **Substanzen, welche auf den TGFβ-Signalweg einwirken**
  - Losartan:
    In Experimenten mit einem Mausmodell für Marfan Syndrom (MFS), einer verwandten Erbkrankheit des Bindegewebes, welche ebenfalls das Risiko für Aneurysmen und Dissektionen erhöht, konnte gezeigt werden, dass der Angiotensin-II-Typ-1-Rezeptor-Antagonist Losartan den Aortenphänotyp reduzieren oder sogar ganz rückgängig machen kann. Dieser positive Effekt kommt dadurch zustande, dass Losartan das TGFβ-Signaling, welches in MFS erhöht ist, reduziert (Habashi et al. 2006, Science 312:117-21).

*Unpublished data*

- **Weitere Substanzen, bei denen eine Wirkung auf das Bindegewebe gezeigt wurde**
  - Celiprolol (β-Blocker):
    Vor kurzem wurde eine klinische Studie mit dem blutdrucksenkenden Mittel Celiprolol abgeschlossen und es konnte gezeigt werden, dass bei Patienten mit EDS IV mit diesem β-Blocker Dissektionen und Rupturen der Arterien verhindert werden können. Bei dieser Studie wurde jedoch nicht untersucht, ob die Substanz einen Effekt auf die Struktur der extrazellulären Matrix der Aortenwand hat (Ong et al. 2010, Lancet 376:1476-84).

*Unpublished data*

**Versuch Teil 2**

Die vielversprechendsten Substanzen aus Teil 1 werden ausgewählt (~3) und es werden weiterführende Tests durchgeführt:

- verschiedene Dosen (2a)
- Therapie und Beobachtung der Tiere über längere Zeiträume (2b)
- Testen der Substanzen beim zweiten, bekannteren Mausmodell für EDS IV (C.129S4(B6)-Col3a1tm1Jae/J) (2c)

51.2 Begründung für die Wahl der Methode (bzw. des Modells) unter Darstellung der Besonderheiten/Vorteile

Die verminderte mechanische Stabilität der Aorta und das erhöhte Risiko für Rupturen sind die schwerwiegendste Komplikation bei EDS IV und verwandten Krankheiten. Deshalb fokussieren wir beim Testen der Substanzen auf deren vaskuläre Wirkung.

Verteiler: Original des Gesuchs und mind. 3 Kopien an die Bewilligungsbehörde

**Gesuch** (Form. A) **- 7 -** Nr. .....
(von der Bewilligungsstelle auszufüllen)

**Kurztitel:** Überprüfung der Wirkung von verschiedenen Substanzen in Mausmodellen für Ehlers-Danlos Syndrom vaskulärer Typ (EDS IV), Kurztitel: Therapieansätze EDS IV Maus

51.3 Begründung für die Wahl der Tierarten (Stämme sind aufzulisten) und falls zutreffend für das Verwenden von Tieren, die nicht zu Versuchszwecken gezüchtet wurden

Bei beiden beschriebenen Mausmodellen für EDS IV entspricht der genetische Defekt mit echter Haploinsuffizienz für *Col3a1* dem unserer Patienten, was sie zu einem guten Modell für unsere Studien macht. Bei den Col3 alpha 1 delta Mäusen entspricht der Phänotyp mit spontanen Todesfällen viel mehr dem der Patienten mit EDS IV, weshalb wir hauptsächlich mit alpha 1 delta arbeiten werden und C.129S4(B6)-Col3a1tm1Jae/J nur für die Bestätigung unserer Resultate verwenden werden.

52 Spezielle Vorbereitung der Tiere auf den Versuch (Angewöhnung, Markierungsart, Konditionierung, Futter- oder Wasserentzug, Vorbehandlung etc.)

Die Tiere werden in der Tierhaltung gezüchtet. Beim Absetzen werden sie mit Ohrstanzen markiert und genotypisiert gemäss Standardprotokoll.

Nach der Zuteilung in die entsprechenden Gruppen haben sie eine Woche Eingewöhnungszeit.

53.1 Anästhesie und/oder weitere Schmerzbekämpfung (Mittel, Dosen, Applikationsweg und -häufigkeit, Zeitdauer etc.)

Keine

53.2 Begründung für die Wahl der Anästhesie oder Analgesie sowie ggf. Begründung für den Verzicht auf diese oder andere belastungsmindernde Massnahmen (z.B. Schmerzmitteleinsatz)

Da es beim neuen Mausmodell (Col3 alpha 1 delta) keine Anzeichen für eine drohende Ruptur der Aorta gibt, können auch keine belastungsmindernden Massnahmen, wie Euthanasierung, angewendet werden. Der Tod durch Aortenruptur wird jedoch schnell eintreten und nicht mit grossem Leiden der Tiere verbunden sein. Ansonsten werden keine Beeinträchtigungen des Wohlbefindens oder Schmerzen erwartet. Auch beim älteren Mausmodell mit weniger ausgeprägtem Phänotyp (C.129S4(B6)-Col3a1tm1Jae/J) werden keine Beeinträchtigungen erwartet. Jegliche Anzeichen von verändertem Verhalten, Schmerzen oder Leiden sind Abbruchkriterien und das Tier wird unverzüglich euthanasiert.

54.1 Art der Eingriffe/Manipulationen und Erheben von Parametern am Tier (Ablaufschema für das einzelne Tier/für die Tiergruppe angeben): operative Eingriffe (Ablauf), Substanzapplikation (Art und Ort, Menge und Häufigkeit), Infizierung, physikalische Einwirkungen (Bestrahlungen etc.), Verlaufskontrollen, Probenerhebung, Reaktionstests etc.

**Substanzen:**

| Substanz | Applikation[a] | Max. Konzentration/Dosis[b] | Häufigkeit |
|---|---|---|---|
| Celiprolol | p.o. | 200mg/kg Körpergewicht pro Tag[c] | durchgehend |
| | | | |
| Doxycyclin | p.o. | 800mg/kg (100mg/kg Körpergewicht/Tag)[e] | durchgehend |
| Losartan | p.o. | 0.6g/l[f] | durchgehend |

*Unpublished data*

[a] p.o.: Die Substanz wird wann immer möglich in Zuckerlösung, mit dem Futter oder Trinkwasser verabreicht und nur falls nicht anders möglich per Gavage.
[b] höchste Dosis, von der beschrieben wurde, dass sie nicht schädlich ist.
[c] Liao et al. 2004, Circulation 110:692-9.

[e] Briest et al. 2011, J Pharmacol Exp Ther 337:621-7.
[f] Habashi et al. 2006, Science 312:117-21.

Verteiler: Original des Gesuchs und mind. 3 Kopien an die Bewilligungsbehörde

**Gesuch** (Form. A) **- 8 -** Nr. .....
(von der Bewilligungsstelle auszufüllen)

**Kurztitel:** Überprüfung der Wirkung von verschiedenen Substanzen in Mausmodellen für Ehlers-Danlos Syndrom vaskulärer Typ (EDS IV), Kurztitel: Therapieansätze EDS IV Maus

**Behandlungsdauer**
Die Behandlungsdauer wird durch den Pilotversuch bestimmt, wird aber maximal 3 Monate betragen. Danach werden die Mäuse mit $CO_2$ euthanasiert und die Aorta wird für Messungen der mechanischen Stabilität, RNA-Extraktion und Histologie verwendet.

**Spontane Todesfälle wegen Aortendissektion**
Die toten Mäuse werden im Versuchsprotokoll festgehalten und aus dem Käfig entfernt und soweit möglich ebenfalls für die Postmortem-Analysen verwendet.

54.2 Dauer des Versuchs/der Versuchsserie: gesamte Versuchsdauer für jede einzelne Gruppe oder jedes Tier, inkl. Zeitdauer, während der das Tier Substanzen oder anderen Noxen ausgesetzt ist

Die Dauer der Versuche hängt für beide Mausmodelle vom Pilotversuch ab, wird aber maximal 3 Monate (+1 Woche Eingewöhnungszeit) betragen. Bei Teil 2b (s. 54.3) wird sie max. 1 Jahr betragen. Bei welchem Alter die Versuche durchgeführt werden, ist ebenfalls vom Pilotversuch abhängig und kann bei Versuch 2c (mit dem zweiten Mausmodell) anders sein als bei den restlichen Versuchen.

54.3 Tiere pro Versuch/Versuchsserie: Anzahl Gruppen (alle Untersuchungsvariablen, z.B. Dosen, Zeitdauer, Kontrollen) und Anzahl Tiere pro Gruppe

Hinweis: Teilweise werden Substanzen auch parallel getestet, wodurch unbehandelte Kontrollgruppen (Wildtyp und Heterozygot) wegfallen werden.

**Pilotversuch**

Der Pilotversuch wird für die zwei Mausmodelle getrennt durchgeführt.

**Col3 alpha1 delta Mäuse**
5 x 2 Gruppen à je 10 Tieren werden beobachtet und zu verschiedenen Zeitpunkten im Alter zwischen 3 Wochen und 4 Monaten euthanasiert, um die mechanische Stabilität der Aorta zu messen.
- Wildtyp
- Heterozygot

**C.129S4(B6)-Col3a1tm1Jae/J Mäuse**
- 2 Gruppen à je 10 Tieren werden in dem Alter, welches beim Pilotversuch mit den Col3 alpha1 delta Mäusen bestimmt wurde, euthanasiert, um die mechanische Stabilität der Aorta zu messen.
  - Wildtyp
  - Heterozygot
- Wenn es beim ersten getesteten Alter keinen signifikanten Unterschied gab, werden noch 4 x 2 Gruppen à je 10 Tieren zu verschiedenen Zeitpunkten zwischen dem schon getesteten Alter und 21 Monaten euthanasiert, um die mechanische Stabilität der Aorta zu messen.
  - Wildtyp
  - Hetrozygot

**Versuch Teil 1** (mit Col3 alpha1 delta Mäusen)
Pro Testsubstanz wird es jeweils 3 Gruppen à je ~20 Tieren geben.
- Wildtyp unbehandelt (Vehikel)
- Heterozygot unbehandelt (Vehikel)
- Heterozygot behandelt

→ Für alle 10 Substanzen

**Versuch Teil 2:**
Jeweils für die 3 besten Substanzen aus Teil 1

**Teil 2a: verschiedene Dosen** (mit Col3 alpha1 delta Mäusen)
Pro Testsubstanz wird es jeweils 5 Gruppen à je 20 Tieren geben
- Wildtyp unbehandelt (Vehikel)
- Heterozygot unbehandelt (Vehikel)
- Heterozygot behandelt Dosis 1
- Heterozygot behandelt Dosis 2
- Heterozygot behandelt Dosis 3

Verteiler: Original des Gesuchs und mind. 3 Kopien an die Bewilligungsbehörde

**Gesuch** (Form. A) **- 9 -** Nr. .....
(von der Bewilligungsstelle auszufüllen)

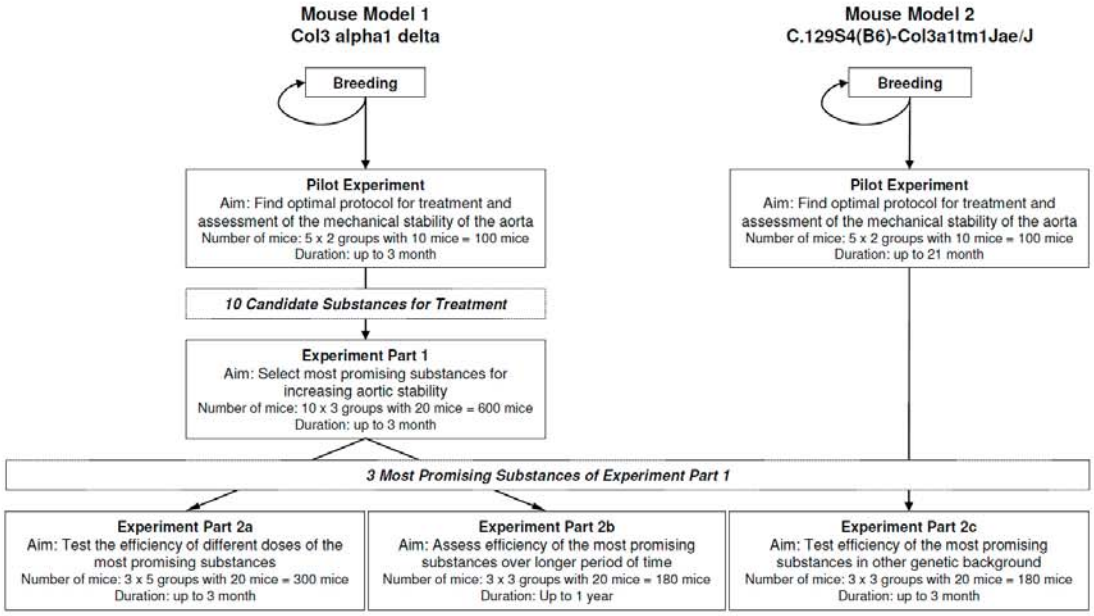
**Kurztitel:** Überprüfung der Wirkung von verschiedenen Substanzen in Mausmodellen für Ehlers-Danlos Syndrom vaskulärer Typ (EDS IV), Kurztitel: Therapieansätze EDS IV Maus

**Teil 2b: längere Behandlungsdauer** (mit Col3 alpha1 delta Mäusen)
Pro Testsubstanz wird es jeweils 3 Gruppen à je ~20 Tieren geben, welche dann max. 1 Jahr behandelt werden.
- Wildtyp unbehandelt (Vehikel)
- Heterozygot unbehandelt (Vehikel)
- Heterozygot behandelt

**Teil 2c: Zweites Mausmodell** (C.129S4(B6)-Col3a1tm1Jae/J Mäusen)
Pro Testsubstanz wird es jeweils 3 Gruppen à je ~20 Tieren geben.
- Wildtyp unbehandelt (Vehikel)
- Heterozygot unbehandelt (Vehikel)
- Heterozygot behandelt

**Zusammenfassung der Versuche und Anzahl benötigter Tiere**

| | Anzahl Experimente | Anzahl Tiere pro Experiment | | | Total Tiere |
|---|---|---|---|---|---|
| | | Wildtyp | Col3 alpha 1 delta / C.129S4(B6)-Col3a1tm1Jae/J | Col3 alpha 1 delta / C.129S4(B6)-Col3a1tm1Jae/J | |
| | | Unbehandelt/Vehikel | Unbehandelt/Vehikel | Substanz | |
| Pilotversuch | 10 | 10 | 10 / 10 | 0 / 0 | 200 |
| Teil 1 | 10 | 20 | 20 / 0 | 20 / 0 | 600 |
| Teil 2a | 3 | 20 | 20 / 0 | 60 / 0 | 300 |
| Teil 2b | 3 | 20 | 20 / 0 | 20 / 0 | 180 |
| Teil 2c | 3 | 20 | 0 / 20 | 0 / 20 | 180 |
| Total | 29 | 480 | 370 / 110 | 440 / 60 | 1460 |




54.4 Begründung für die vorgesehenen Tierzahlen (obligatorische Angabe der eingesetzten Stämme)

Für diesen Versuch ist es wichtig, Unterschiede in der mechanischen Stabilität feststellen zu können. Die Daten von Cooper et al. 2010 (Vet Pathol 47:1028) zeigen jedoch, dass die Streuung relativ gross sein kann. Aufgrund der Sterberate von 28% (Konfidenzintervall 22,5%-34,2%) braucht es eine höhere Anzahl Tiere, damit es von jeder Gruppe genug überlebende Tiere gibt, um einen aussagekräftigen t-Test durchführen zu können. Die Gruppengrössen entsprechen der Minimalgrösse, bei welchen noch statistisch gut auswertbare Werte zu erwarten sind.

Verteiler: Original des Gesuchs und mind. 3 Kopien an die Bewilligungsbehörde

**Gesuch** (Form. A) **- 10 -** Nr. .....
(von der Bewilligungsstelle auszufüllen)

**Kurztitel:** Überprüfung der Wirkung von verschiedenen Substanzen in Mausmodellen für Ehlers-Danlos Syndrom vaskulärer Typ (EDS IV), Kurztitel: Therapieansätze EDS IV Maus

55 Art der Auswertung einschliesslich biometrische Planung

Postmortem wird die mechanische Stabilität der Aorta untersucht. Dazu wird gemessen bei welchem Druck die Aorta rupturiert. Je höher der Druck bei der Ruptur ist, desto stabiler ist die Aorta. Ausserdem wird auch noch die Mortalitätsrate betrachtet.

Für die erfassten Parameter wird jeweils das 95%-Konfidenzintervall angegeben.

- Skalierbare Werte: Mittelwert ± $t_{crit}$*SE (basierend auf Student t-Test; Sokal RR, Rohlf FJ. 1995. Biometry: the principles and practice of statistics in biological research. New York: W. H. Freeman and Co. p 143.152.)
- Relative Häufigkeiten: http://faculty.vassar.edu/lowry/prop1.html

56.1 Erwartete Auswirkungen auf das Befinden der Tiere (Aktivität, Futter- und Wasseraufnahme, Schmerzreaktionen, Dauer und Verlauf der Beeinträchtigung, weitere Verhaltensparameter, Wachstum, erwartete Todesfälle.)

Bei den heterozygoten Col3 alpha 1 delta Mäusen ist eine spontane Mortalitätsrate von ~28% beschrieben, wobei die meisten Todesfälle im Alter von 4-10 Wochen erwartet werden. Die Todesursache ist eine Aortendissektion, welche zu einer Ruptur der Aorta führt. Weitere Beeinträchtigungen sind bis zu einem Alter von 16 Wochen nicht zu erwarten. Für ältere Tiere gibt es bis jetzt noch keine Daten.

Bei den heterozygoten C.129S4(B6)-Col3a1tm1Jae/J Mäusen sind keine Auswirkungen auf das Befinden zu erwarten.

Aufgrund der verabreichten Substanzen werden keine Beeinträchtigungen erwartet, da alle Substanzen schon in Tierversuchen mit Mäusen verwendet wurden und keine Nebenwirkungen beschrieben sind. Ausserdem wurden die meisten Substanzen auch schon beim Menschen eingesetzt und sie sind in den meisten Fällen auch schon zugelassen.

56.2 Überwachung des Befindens: Häufigkeit, Beurteilungskriterien, Dokumentation (z.B. score sheet) entsprechend der Versuchsphase

Die Käfige werden mindestens 5x pro Woche kontrolliert und tote Tiere werden unverzüglich aus den Käfigen entfernt. Der spontane Tod durch Aortenruptur erfolgt schnell und ohne Vorzeichen, da die Tiere nur ganz kurz, wenn überhaupt, leiden werden. Es gibt deshalb keine Möglichkeit, den Tod durch Aortenruptur frühzeitig zu erkennen und zu verhindern. Um dies zu bestätigen, werden in der Anfangsphase der Zucht (in der heterozygote Tiere spontan sterben können) die Tiere über einige Wochen 2-3x pro Woche auf ihr Wohlbefinden untersucht, indem Parameter wie Verhalten, Bewegung, Gewicht und Temperatur kontrolliert werden (siehe Score Sheet). Anhand der Resultate dieser Untersuchungen wird dann festgelegt, ob und wie häufig und umfassend die Tiere während dem Versuch kontrolliert werden müssen. Beim zweiten Mausmodell (C.129S4(B6)-Col3a1tm1Jae/J), welches kein beeinträchtigtes Befinden zeigt, werden die Mäuse während dem Versuch gleich kontrolliert, wie die Col3 alpha1 delta Mäuse.

56.3 Kriterien für belastungsmindernde Massnahmen und (vorzeitigen) Versuchsabbruch (Abbruchkriterien) und für Verzicht auf Wiederverwertung,

Belastungsmindernde Massnahmen sind nicht möglich, da die einzige erwartete Beeinträchtigung, nämlich der spontane Tod durch Aortenruptur nicht hervorgesehen und somit auch nicht verhindert werden kann. Werden jedoch bei einem Tier dennoch Hinweise auf eine Belastung festgestellt, wird es vom Versuch ausgeschlossen und mit $CO_2$ euthanasiert. Die genauen Abbruchkriterien werden aufgrund der Beobachtungen während der Anfangsphase der Zucht (siehe 56.2) festgelegt. Sollte es entgegen der Angaben in der Literatur dennoch Anzeichen für eine bevorstehende Ruptur der Aorta geben, wäre dies dann ein Abbruchkriterium und die Mäuse würden umgehend mit $CO_2$ euthanasiert werden. Beim zweiten Mausmodell werden dann die gleichen Abbruchkriterien angewendet, auch wenn gemäss Literaturangaben keine Belastungen zu erwarten sind.

56.4 Geschätzte Anzahl Tiere pro Schweregrad
Hinweis: Teilweise werden Substanzen auch parallel getestet, wodurch unbehandelte Kontrollgruppen (Wildtyp und Heterozygot) wegfallen werden.

Col3 alpha 1 delta Wildtyp: 370
- Behandlung mit Vehikel → SG 1: 320

Col3 alpha 1 delta Heterozygot: 810
- Risiko für Tod durch Aortendissektion (~28%) → SG 3: 810
- Behandlung mit Vehikel → SG 1: 320
- Behandlung mit den Substanzen → SG 1:440

Verteiler: Original des Gesuchs und mind. 3 Kopien an die Bewilligungsbehörde

**Gesuch** (Form. A) **- 11 -** Nr. .....
(von der Bewilligungsstelle auszufüllen)

**Kurztitel:** Überprüfung der Wirkung von verschiedenen Substanzen in Mausmodellen für Ehlers-Danlos Syndrom vaskulärer Typ (EDS IV), Kurztitel: Therapieansätze EDS IV Maus

C.129S4(B6)-Col3a1tm1Jae/J Wildtyp: 110
- Behandlung mit Vehikel → SG 1: 60

C.129S4(B6)-Col3a1tm1Jae/J Heterozygot: 170
- Leichte Belastung durch den Phänotyp → SG 1: 170
- Behandlung mit Vehikel → SG 1: 60
- Behandlung mit den Substanzen → SG 1: 60

57.1 Haltung und Pflege der Tiere vor, während, zwischen und nach Einzelversuchen (Platzangebot, Strukturierung, Auslauf, Einzel- oder Gruppenhaltung, Fütterung und Beschäftigung, Routinekontrollen durch Tierpfleger/innen; bei wiederholter Verwendung auch Abstand zwischen den Versuchen)

Die Mäuse werden in Gruppenhaltung in Standardkäfigen mit Enrichement gehalten. Futter und Wasser werden ad libitum zur Verfügung gestellt. Die Futteraufnahme und das Wohlbefinden werden regelmässig überprüft.

57.2 Begründung für allfällige Abweichungen von den Haltungsbedingungen gemäss Tierschutzverordnung (inkl. Futterentzug, Immobilisation)

N/A

58 **Tötungsmethode**, Verwendung der Tiere nach Abschluss des (Einzel-) Versuchs (wiederholter Einsatz im gleichen bzw. in anderem Versuch)

Die Mäuse werden mit $CO_2$ euthanasiert

## 6 ANGABEN ZUR BEGRÜNDUNG DES TIERVERSUCHS

61 Welche anderen Versuchsmethoden sind (z.B. aus der Literatur) bekannt, die es ermöglichen, entsprechende Information zu erhalten (In-vitro oder In-vivo Methoden angeben)

Es sind keine weniger invasive oder in vitro Methoden bekannt, um den Effekt der Substanzen auf die mechanische Stabilität der Aorta und das Risiko für Dissektionen so gut zu bestimmen, wie mit den beschriebenen Versuchsmethoden.

62 Angabe, ob das Projekt begutachtet wurde/wird, und wenn ja, von welcher Institution/Organisation

Die vorgeschlagenen Experimente sind Teil des SNF-Fortsetzungsprojektes von PD Dr. Gabor Matyas (Leiter der Dissertation von Janine Meienberg).

63 Beurteilung der Bedeutung des erwarteten Erkenntnisgewinns oder Ergebnisses im Vergleich zu den den Tieren entstehenden Schmerzen, Leiden, Schäden oder Ängsten

Der zu erwartende spontane Tod durch Ruptur der Aorta bei heterozygoten Col3 alpha 1 delta Mäusen, tritt schnell ein und ist mit wenig Leiden verbunden. Somit werden die Tiere trotz der grossen Belastung durch den plötzlichen Tod, verhältnismässig wenig, und nicht mehr als natürlicherweise/sonst der Fall wäre, leiden. Der Schweregrad, welcher nach Rücksprache mit dem kantonalen Veterinäramt (Claudia Lawnitzak) auf 3 festgelegt wurde, ist daher eher hoch eingestuft. Alle heterozygoten Tiere tragen das Risiko für diese Belastung, aber nur ein Teil davon wird sie wirklich erfahren. Die Zahl der Tiere mit der schweren Schädigung wird deshalb retrospektiv deutlich kleiner sein, als prospektiv veranschlagt. Ausserdem werden durch Staging der Versuche (siehe 56.4) noch heterozygote Tiere eingespart werden können, wodurch die Zahl der belasteten Tiere nochmals gesenkt werden kann.

Die zusätzliche Belastung der Tiere durch den Versuch wird sehr klein sein, da keine invasiven Eingriffe geplant sind, ausser der Injektion von einigen der Testsubstanzen. Die meisten dieser Substanzen werden übers Futter oder Trinkwasser appliziert. Ausserdem sind bei allen Substanzen in den gewählten Dosierungen keine Nebenwirkungen zu erwarten.

Dem gegenüber steht, dass die Identifikation einer Substanz, welche die Stabilität der Aorta erhöht und somit die Mortalitätsrate verringert, für die EDS-IV-Patienten von grosser Bedeutung ist, wenn so das Risiko für eine lebensbedrohliche Situation gesenkt werden kann. Das erhöhte Risiko für Tod durch Aortenruptur ist in Familien mit genetisch bedingtem erhöhtem Risiko für Aortendissektionen eine grosse Belastung. Da diese meist sehr plötzlich auftritt, können auch regelmässige Kontrollen das Risiko nur bedingt reduzieren. Der Lebensstil muss angepasst werden und gewisse Berufe und Sportarten sollten nicht ausgeführt werden. Dies führt zu einer Verminderung der Lebensqualität. Dazu kommt, dass in betroffenen Familien oft schon Familienmitglieder verstorben sind, was die Belastung zusätzlich erhöht.

Verteiler: Original des Gesuchs und mind. 3 Kopien an die Bewilligungsbehörde

**Gesuch** (Form. A) **- 12 -** Nr. .....
(von der Bewilligungsstelle auszufüllen)

**Kurztitel:** Überprüfung der Wirkung von verschiedenen Substanzen in Mausmodellen für Ehlers-Danlos Syndrom vaskulärer Typ (EDS IV), Kurztitel: Therapieansätze EDS IV Maus

Die Resultate von diesem Versuch dürften direkt auf Fälle mit echter Haploinsuffizienz für *COL3A1*, wie die von uns beschriebene Familie, angewendet werden, da es sich um den gleichen Gendefekt wie in den Mausmodellen handelt. Die Resultate dürften aber auch allen anderen Patienten mit EDS IV helfen, bei denen zu wenig funktionierendes Kollagen vorliegt.

Darüber hinaus gibt es die Möglichkeit aus den Erkenntnissen dieser Studie Nutzen für verwandte Krankheiten, welche ebenfalls zu Dissektionen und Rupturen in der Aorta führen können, zu ziehen. Da wir hauptsächlich Substanzen testen, die für den Menschen schon zugelassen sind, können die Resultate direkt und sehr schnell bei Patienten angewendet werden und hoffentlich deren Perspektive und Lebensqualität verbessern.

Fazit: Wir sind der Ansicht, dass der Nutzen der Versuche die Belastung der Tiere klar überwiegt und halten deshalb diese Versuche für gerechtfertigt. Ausserdem haben wir uns bemüht, die Zahl der verwendeten Tiere so tief wie möglich zu halten. Die Experimente wären mit weniger Tieren wissenschaftlich nicht interpretierbar.

Verteiler: Original des Gesuchs und mind. 3 Kopien an die Bewilligungsbehörde

# Appendix 6 Poster: Assessment of the Mechanical Stability of the Aorta in a *Col3a1* Mouse Model

Muenger J, Meienberg J, Crabb J, Mauri A, Gysi S, Kaiser C, Barmettler G, de Vos J, Bhattacharya I, Courseau J, Giunta C, Bakker EN, Battegay EJ, Jaeger R, van Bavel E, Haas E, Ziegler U, Kopf M, Zeisberger S, Mazza E, Matyas G (2015) Assessment of the mechanical stability of the aorta in a Col3a1 mouse model. 14th Day of Clinical Research (DCR), Zurich, April 9, 2015.

Zentrum für Kardiovaskuläre Genetik und Gendiagnostik
Das Genetikzentrum der Stiftung für Menschen mit seltenen Krankheiten

ZIHP

University of Zurich[UZH]

# Assessment of the Mechanical Stability of the Aorta in a *Col3a1* Mouse Model

Münger J,[1] Meienberg J,[1] Crabb J,[2] Mauri A,[2] Gysi S,[1] Kaiser C,[3] Barmettler G,[3] De Vos J,[4] Bhattacharya I,[5] Courseau J,[6] Giunta C,[7] Bakker E,[4] Battegay E,[5] Jaeger R,[6] Van Bavel E,[4] Haas E,[5] Ziegler U,[3] Zeisberger S,[8] Mazza E,[2] Matyas G[1,9]

[1]Center for Cardiovascular Genetics and Gene Diagnostics, Foundation for People with Rare Diseases, Schlieren-Zurich, Switzerland, [2]Institute of Mechanical Systems, Swiss Federal Institute of Technology Zurich, Zurich, Switzerland, [3]Center for Microscopy and Image Analysis, University of Zurich, Zurich, Switzerland, [4]Department of Biomedical Engineering and Physics, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands, [5]Research Unit, Division of Internal Medicine, University Hospital of Zurich, Zurich, Switzerland, [6]Fraunhofer Institute for Mechanics of Materials IWM, Freiburg, Germany, [7]Division of Metabolism, University Children's Hospital, Zurich, Switzerland, [8]Swiss Centre for Regenerative Medicine, University of Zurich, Zurich, Switzerland, [9]Zurich Center for Integrative Human Physiology, University of Zurich, Zurich, Switzerland

Contacts: matoga@genetikzentrum.ch, mazza@imes.mavt.ethz.ch, matyas@genetikzentrum.ch

## Conclusions

- **Our data show significant differences between ascending and descending thoracic aorta as well as between heterozygous and wild-type *Col3a1* mice.**
- **Our protocol is suitable for the assessment of the mechanical stability of mice aorta, opening the way to test pharmacological substances for their potential to increase the mechanical stability of the aorta with the goal to find a targeted therapy that lowers the risk for aortic ruptures in EDS IV patients.**

## Introduction

Ehlers-Danlos syndrome vascular type (EDS IV) is a rare connective tissue disorder (~2 in 100'000) characterized by translucent skin, easy bruising, and arterial, intestinal and/or uterine fragility. The most severe complication is the increased risk for rupture of the aorta, leading to life-threatening internal bleeding. EDS IV is inherited in an autosomal dominant manner and caused by mutations in the *COL3A1* gene, which encodes the alpha 1 chain of type III collagen, a fibrillar collagen expressed in walls of hollow organs. So far, only disease management and treatment of symptoms are available but no targeted therapy. Recently, a novel mouse model which has true haploinsufficiency of *Col3a1* due to a spontaneous deletion has been described [Smith *et al.* 2011, Cardiovasc Res 90:182-190]. In heterozygous mice, this *Col3a1* deletion leads to reduced mechanical stability of the aorta and, in ~28% of cases, to spontaneous rupture of the aorta and thus to increased mortality similar to the phenotype of human patients. Our goal was to determine whether in this mouse model significant difference in mechanical stability of the thoracic aorta between heterozygous and wild-type animals can be measured and hence objectively characterized. This could allow us in the future to find substance(s) that will increase the mechanical stability of the aortic wall and therefore reduce mortality due to aortic dissection.

## Methods and Results

The thoracic aorta from wild-type and heterozygous *Col3a1* mice was dissected and cleaned of adherent connective tissue. 1.5-mm-long sections of aortic arch as well as ascending and descending aorta (Figure 1) were mounted on two 200-µm diameter stainless steel wires on a Tissue Puller (Danish Myo Technology). The mounted vessels were stretched radially until tissue damage, thereby recording the stretching force (in mM). Maximum force at tissue damage was significantly lower in heterozygous *Col3a1* mice compared to age- and gender-matched wild-type animals in both the ascending and descending parts of the aorta. For both genotypes, the mechanical stability of the aorta was decreasing with increasing distance from the heart (Figures 2 and 3). Collagen distribution throughout the aortic wall and the orientation of collagen fibers was imaged by second harmonic generation (SHG) from nonlinear laser scanning microscopy (Figure 4). In the aortic samples of heterozygous mice, the analysis of the collagen content from the SHG microscopy images in the unstretched configuration showed a significantly smaller collagen volume. In addition, the density coefficient was significantly smaller in samples of heterozygous compared to wild-type mice, indicating larger spaces among collagen fibers (Table 1). Analyzing the images of the collagen network during stretching revealed further differences between aortic samples of heterozygous and wild-type mice (Figure 5). While in the tissue of wild-type animal the collagen fibers started reorientating and aligning at first stretching step (1.9 mm), tissue of heterozygous mice showed little reaction until the last stretching step (3.2 mm). This is supported by the trend of decreasing collagen volume during stretching for the tissue of wild-type mice, indicating an alignment of the collagen fibers leading to a densely packed network. The reduced alignment and orientation might offer an explanation for the reduced stiffness of the arterial tissue of the heterozygous mice as the collagen network is no longer able to resist high loads. Reduced collagen content and larger spaces among collagen fibers of aortic sections of heterozygous animals where confirmed by electron microscopy images (Figure 6).

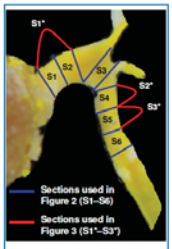



Figure 1. Location of 1.5-mm-long sections of the thoracic aorta used for measurements.

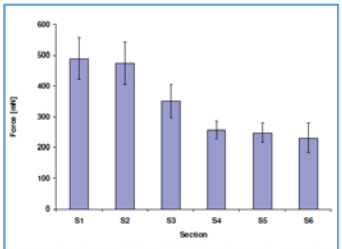



Figure 2. Mechanical properties of the sections S1-S6 (see Figure 1) of mouse aortas from 15-month-old wild-type females (n = 10), determined in a Tissue Puller setup. Error bars indicate 95% confidence intervals.

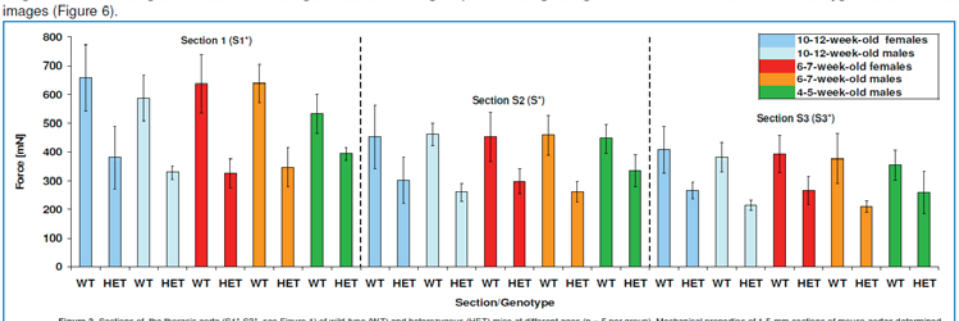



Figure 3. Sections of the thoracic aorta (S1*-S3*, see Figure 1) of wild-type (WT) and heterozygous (HET) mice at different ages (n = 5 per group). Mechanical properties of 1.5-mm sections of mouse aortas determined in a Tissue Puller setup. Section S1*: aorta ascendens, sections S2* and S3*: aorta descendens. Error bars indicate 95% confidence intervals.

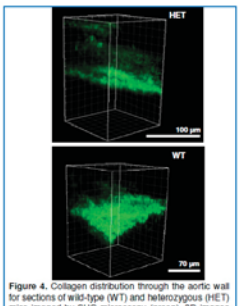



Figure 4. Collagen distribution through the aortic wall for sections of wild-type (WT) and heterozygous (HET) mice imaged by SHG microscopy (green). 3D images of the whole wall thickness.

| | Sample | Unstretched | λ = 1.9 mm | λ = 2.6 mm | Density coefficient |
|---|---|---|---|---|---|
| WT | Mean | 1'443'702 | 936'737 | - | 0.132 |
| | STD | 191'936 | 168'743 | - | 0.029 |
| | M135 | 1'493'695 | 860'823 | 502'990 | 0.133 |
| | M240 | 1'231'716 | 819'286 | 495'572 | 0.103 |
| | M241 | 1'605'694 | 1'130'103 | - | 0.160 |
| HET | Mean | 981'570 | 793'458 | - | 0.069 |
| | STD | 36'872 | 56'807 | - | 0.007 |
| | M134 | 940'820 | 151'781 | - | 0.077 |
| | M6 | 1'012'627 | 1'235'659 | - | 0.069 |
| | M239 | 991'264 | 992'934 | 638'724 | 0.063 |
| | p-value | 0.015 | NS | - | 0.021 |

Table 1. Collagen volume and density calculated at various stretches of aortic sections of wild-type (WT) and heterozygous (HET) mice. Volume values are given in µm³ and mean and standard deviation (STD) are noted. In the unstretched configuration, a significant difference in volume ($p = 0.015$) and density ($p = 0.021$) is found between samples of heterozygous and wild-type animals.

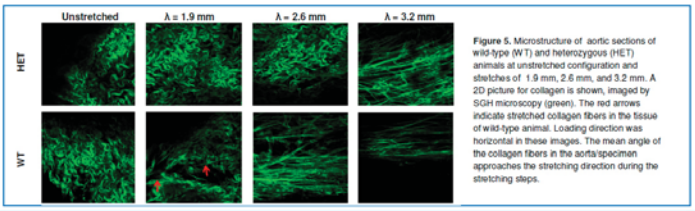



Figure 5. Microstructure of aortic sections of wild-type (WT) and heterozygous (HET) animals at unstretched configuration and stretches of 1.9 mm, 2.6 mm, and 3.2 mm. A 2D picture for collagen is shown, imaged by SGH microscopy (green). The red arrows indicate stretched collagen fibers in the tissue of wild-type animal. Loading direction was horizontal in these images. The mean angle of the collagen fibers in the aorta/specimen approaches the stretching direction during the stretching steps.

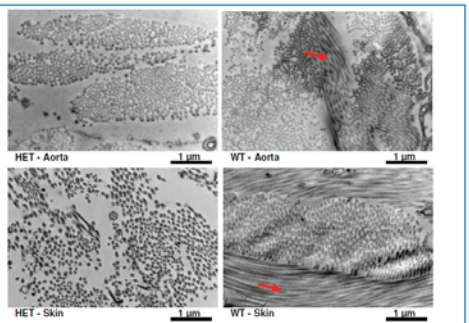



Figure 6. Electron microscopy images of thoracic aorta and abdominal skin sections of 16-week-old heterozygous (HET) and wild-type (WT) animals. The red arrows indicate longitudinal aligned collagen fibers which are only displayed in wild-type tissue. Microscope magnification: 24'500×.

**Acknowledgement:** This work was supported by grants from the Gebauer Stiftung, Isaac Dreyfus-Bernheim Stiftung, Parrotia-Stiftung, Sutter-Stöttner Stiftung, and Wolfermann-Nägeli-Stiftung.

## Contribution of Authors

| | |
|---|---|
| Justyna Muenger | Tissue Puller experiments, electron microscopy, data analysis, writing and editing of the poster |
| **Janine Meienberg** | **Conceiving and planning the study, data analysis, writing and editing of the poster** |
| Jessica Crabb | Stretching experiments and multiphoton microscopy, writing and editing of the poster |
| Arabella Mauri | Stretching experiments and multiphoton microscopy, editing of the poster |
| Carmen Kaiser | Electron microscopy, editing of the poster |
| Gery Barmettler | Electron microscopy, editing of the poster |
| Cecilia Giunta | Contribution to data analysis and editing of the poster |
| Urs Ziegler | Electron microscopy, editing of the poster |
| Steffen Zeisberger | Conceiving and planning the study, editing of the poster |
| Edoardo Mazza | Conceiving and planning the study, editing of the poster |
| Gabor Matyas | Initiation of the study, conceiving and planning the study, writing and editing of the poster |

# Appendix 7 Curriculum Vitae


| | |
|---|---|
| Name | MEIENBERG |
| First name | Janine |
| Academic title | M.Sc. |
| Date of birth | March 19, 1985 (place of birth: Neuheim ZG, Switzerland) |
| Marital status | unmarried |
| Address | Center for Cardiovascular Genetics and Gene Diagnostics<br>Wagistrasse 25, CH-8952 Schlieren<br>phone +41 43 433 86 75<br>e-mail meienberg@genetikzentrum.ch |
| | Private: Wolfetslohstrasse 1<br>CH-8907 Wettswil<br>+41 44 700 02 56<br>j_meienberg@yahoo.de |


**EDUCATION**

- Since 2010 PhD studies, University of Zurich, Switzerland
- 2007-2009 M.Sc. in Human Biology (summa cum laude), University of Zurich, Switzerland
- 2004-2007 B.Sc. in Biology (summa cum laude), University of Zurich, Switzerland
- 1998-2004 Matura (MAR; magna cum laude), Kantonsschulen Freudenberg & Wiedikon, Zurich, Switzerland

**AWARDS**

- Mai 2009 Award for Master Thesis from the University of Zurich, Switzerland

**PROFESSIONAL EXPERIENCE**

- **PhD Student**, Center for Cardiovascular Genetics and Gene Diagnostics (2012-present) / University of Zurich, Institute of Medical Molecular Genetics (2010-2011) and Zurich Center for Integrative Human Physiology (ZIHP) (PD Dr.sc.nat. G. Matyas), since May 2010:
  Molecular Basis of Aortic Diseases
- **Research Assistant**, University of Zurich, Institute of Medical Genetics (PD Dr.sc.nat. G. Matyas), Schwerzenbach, Switzerland, March 2009 - December 2009:
  Comprehensive mutation screening for several genes associated with Marfan syndrome and related disorders, assistance in supervision of student trainees
- **Master Student**, University of Zurich (Prof. T. Hennet), Institute of Medical Genetics (PD Dr.sc.nat. G. Matyas), Schwerzenbach, Switzerland, November 2007 - February 2009:
  Assessment of the Role of *COL3A1* Gene Mutations in Patients with Suspected Marfan Syndrome

**APPLIED TECHNIQUES**

Primer design, DNA and RNA extraction from blood and fibroblasts, PCR, long-range PCR, reverse transcription PCR, quantitative real-time PCR using both gDNA and RNA templates, qualitative and semi-quantitative DNA sequencing, MLPA, cell culture, microarray (SNP and aCGH) analyses, library preparation as well as data analysis and evaluation for NGS, and LTK1

**RESEARCH INTEREST**

Molecular basis of Marfan syndrome and Ehlers-Danlos syndrome vascular type with focus on the role of the *COL3A1* gene in the pathogenesis of these and related disorders

**LANGUAGES**

Swiss German (native language), German (excellent knowledge), English (advanced knowledge, visiting a language school in Brighton and some research laboratories in the UK from January to April 2010), French (good knowledge)

**IT-KNOWLEDGE**

MS Word (advanced knowledge), MS Excel (good knowledge), MS PowerPoint Excel (good knowledge), ABI SeqScape (advanced knowledge), BioDiscovery Nexus Copy Number (good knowledge), Linux (basic knowledge)

# Appendix 8 List of Publications

### ORIGINAL ARTICLES WITH PEER REVIEW

1. **Meienberg J***, Zerjavic K*, Keller I, Okoniewski M, Patrignani A, Ludin K, Xu Z, Steinmann B, Carrel T, Roethlisberger B, Schlapbach R, Bruggmann R, Matyas G (2015) New insights into the performance of human whole-exome capture platforms. Nucleic Acids Res 43:e76.
   *equally contributing first authors
2. Okoniewski M*, **Meienberg J***, Patrignani A, Szabelska A, Matyas G, Schlapbach R (2013) Precise breakpoint localization of large genomic deletions using PacBio and Illumina next-generation sequencers. Biotechniques 54:98-100.
   *equally contributing first authors
3. **Meienberg J**, Rohrbach M, Neuenschwander S, Spanaus K, Giunta C, Alonso S, Arnold E, Henggeler C, Regenass S, Patrignani A, Azzarello-Burri S, Steiner B, Nygren AOH, Carrel T, Steinmann B, Matyas G (2010) Hemizygous deletion of *COL3A1*, *COL5A2*, and *MSTN* causes a complex phenotype with aortic dissection: a lesson for and from true haploinsufficiency. Eur J Hum Genet 18:1315-1321.

### MEETING ABSTRACTS AND CONFERENCE PROCEEDINGS

1. Muenger J, **Meienberg J**, Crabb J, Mauri A, Kaiser C, Barmettler G, Giunta C, Ziegler U, Zeisberger S, Mazza E, Matyas G (2015) Assessment of the mechanical stability of the aorta in a mouse model of Ehlers-Danlos syndrome vascular type (EDS IV). ESHG 2015, *Eur J Hum Genet* 23, Suppl. 1:119..
2. Muenger J, **Meienberg J**, Crabb J, Mauri A, Gysi S, Kaiser C, Barmettler G, de Vos J, Bhattacharya I, Courseau J, Giunta C, Bakker EN, Battegay EJ, Jaeger R, van Bavel E, Haas E, Ziegler U, Kopf M, Zeisberger S, Mazza E, Matyas G (2015) Assessment of the mechanical stability of the aorta in a *Col3a1* mouse model. 14th Day of Clinical Research (DCR), Zurich, April 9, 2015.
3. Muenger J, **Meienberg J**, Crabb J, Mauri A, Kaiser C, Barmettler G, Giunta C, Ziegler U, Zeisberger S, Mazza E, Matyas G (2015) Bestimmung der mechanischen Stabilität der Aorta in einem Mausmodell für Ehlers-Danlos Syndrom vaskulärer Typ (EDS IV). 29. Jahrestagung der Arbeitsgemeinschaft für Pädiatrische Stoffwechsel-Störungen, Fulda, Germany, March 4-6, 2015, *Kinderheilkunde* (in press).
4. **Meienberg J***, Zerjavic K*, Okoniewski M, Patrignani A, Ludin K, Bruggmann R, Xu Z, Steinmann B, Carrel T, Roethlisberger B, Schlapbach R, Matyas G (2014) Evaluation of whole-exome enrichment platforms for genetic testing of aortic diseases. 9th International Research Symposium on Marfan Syndrome and Related Disorders, Paris, France, September 25-27, 2014. *equally contributing first authors
5. Zerjavic K*, **Meienberg J***, Okoniewski M, Patrignani A, Ludin K, Gysi S, Steinmann B, Carrel T, Roethlisberger B, Schlapbach R, Matyas G (2014) Evaluation of three sequence capture platforms for whole-exome sequencing. ESHG 2014, *Eur J Hum Genet* 22, Suppl. 1:282.
   *equally contributing first authors
6. **Meienberg J***, Zerjavic K*, Okoniewski M, Patrignani A, Ludin K, Bruggmann R, Xu Z, Roethlisberger B, Schlapbach R, Matyas G (2014) Comparison of three recent sequence capture platforms for whole-exome sequencing. Human Genome Meeting 2014, Geneva, April 27-30, 2014
   *equally contributing first authors
7. **Meienberg J**, Okoniewski M, Patrignani A, Szabelska A, Tsai Y-C, Carrel T, Steinmann B, Turner SW, Korlach J, Roethlisberger B, Schlapbach R, Matyas G (2014) True haploinsufficiency in rare aortic diseases: Identification and Characterization of large Deletions using Next-Generation Sequencing. RE(ACT) 2nd International Congress on Research of Rare and Orphan Diseases, Basel, March 5-8, 2014.
8. Zerjavic K, **Meienberg J**, Matyas G (2014) Comparison of SureSelect, NimbleGen and Nextera capture platforms for whole-exome sequencing. RE(ACT) 2nd International Congress on Research of Rare and Orphan Diseases, Basel, March 5-8, 2014.

9. **Meienberg J***, Okoniewski M*, Patrignani A, Szabelska A, Tsai YC, Turner SW, Korlach J, Schlapbach R, Matyas G (2013) True haploinsufficiency in aortic diseases: breakpoint characterization of large deletions using Next-Generation Sequencing. 9th Symposium of the Zurich Center for Integrative Human Physiology, Zurich, August 23, 2013.
*equally contributing first authors
10. Venier A, **Meienberg J**, Keshavan R, Okoniewski M, Patrignani A, Shams S, Culot L, O'Hara A, Che Z, Matyas G, Roethlisberger B (2013) Integration of copy number and sequence variation data: an investigation strategy for the molecular basis of aortic diseases. 9th European Cytogenetics Conference, Dublin, June 29- July 2, 2013.
11. **Meienberg J***, Okoniewski M*, Patrignani A, Szabelska A, Tsai YC, Turner SW, Korlach J, Schlapbach R, Matyas G (2013) Breakpoint characterization of large deletions using PacBio and Illumina sequencing technologies. ESHG 2013, *Eur J Hum Genet* 21, Suppl. 1:360.
*equally contributing first authors
12. **Meienberg J**, Okoniewski M, Patrignani A, Tsai YC, Alonso S, Arnold E, Henggeler C, Carrel T, Steinmann B, Korlach J, Turner SW, Schlapbach R, Matyas G (2012) Novel approach for characterization of large deletions using next-generation sequencing. New Frontiers Symposium on Personal Genomics, Nijmegen, December 3-4, 2012.
13. **Meienberg J**, Rohrbach M, Neuenschwander S, Giunta C, Alonso S, Henggeler C, Carrel T, Steinmann B, Matyas G (2012) True haploinsufficiency of *COL3A1* due to hemizygous deletion causes aortic dissection. First International Symposium on the Ehlers-Danlos Syndrome, Ghent, September 8-11, 2012.
14. **Meienberg J**, Alonso S, Patrignani A, Okoniewski M, Arnold E, Henggeler C, Carrel T, Steinmann B, and Matyas G (2012) Characterisation of large hemizygous *FBN1* deletions causing Marfan syndrome. 7th European Elastin Meeting, Ghent, September 1-4, 2012.
15. **Meienberg J**, Patrignani A, Okoniewski M, Henggeler C, Arnold E, Perez R, Mahlberg N, Amstutz N, Burri H, Dutly F, Carrel T, Steinmann B, Matyas G (2012) Evaluation of exome sequencing in genes associated with aortic connective tissue disorders. XXIIIrd Meeting of the Federation of European Connective Tissue Societies, Katowice, August 25-29, 2012.
16. **Meienberg J**, Patrignani A, Okoniewski M, Henggeler C, Arnold E, Perez R, Mahlberg N, Amstutz N, Burri H, Dutly F, Carrel T, Steinmann B, Matyas G (2012) Evaluation of exome sequencing in genes associated with aortic diseases. ESHG 2012, Eur J Hum Genet 20, Suppl. 1:306.
17. **Meienberg J**, Rohrbach M, Neuenschwander S, Spanaus K, Giunta C, Alonso S, Arnold E, Henggeler C, Perez R, Regenass S, Azzarello-Burri S, Carrel T, Steinmann B, Matyas G (2012) Hemizygous deletion leading to true haploinsufficiency of *COL3A1* causes aortic dissection. RE(ACT) International Congress on Research on Rare and Orphan Diseases, Mol Syndromol 2011;2:272.
18. **Meienberg J**, Patrignani A, Okoniewski M, Henggeler C, Arnold E, Perez R, Mahlberg N, Amstutz N, Burri H, Dutly F, Carrel T, Steinmann B, Matyas G (2012) Evaluation of exome sequencing with different types of sequence variations in genes associated with aortic diseases. RE(ACT) International Congress on Research on Rare and Orphan Diseases, Mol Syndromol 2011;2:271.
19. **Meienberg J**, Rohrbach M, Neuenschwander S, Spanaus K, Giunta C, Alonso S, Arnold E, Henggeler C, Perez R, Regenass S, Azzarello-Burri S, Carrel Thierry, Berger W, Steinmann B, Matyas G (2010) Hemizygous deletion comprising *COL3A1* and *COL5A2* causes aortic dissection. 8th International Research Symposium on Marfan Syndrome, Warrenton, September 11-14, 2010.
20. **Meienberg J**, Neuenschwander S, Rohrbach M, Giunta C, Alonso S, Arnold E, Henggeler C, Perez R, Spanaus K, Regenass S, Azzarello-Burri S, Berger W, Steinmann B, Matyas G (2010) True haploinsufficiency of *COL3A1* causes aortic dissection. XXIInd Meeting of the Federation of European Connective Tissue Societies, Davos, July 3-7, 2010 (poster and oral presentation).
21. Spanaus K, **Meienberg J**, Neuenschwander S, Alonso S, Henggeler C, Azzarello-Burri S, Steiner B, Berger W, Matyas G (2010) No evidence for hemochromatosis type 4 in hemizygous *SLC40A1* deletion carriers. ESHG 2010, Eur J Hum Genet 18, Suppl. 1:343.
22. Spanaus K, **Meienberg J**, Neuenschwander S, Alonso S, Henggeler C, Azzarello-Burri S, Steiner B, Berger W, Matyas G (2009) Heterozygous deletion of the *SLC40A1* gene is not associated with type 4 hemochromatosis. Annual meeting of the Swiss Society of Clinical Chemistry, Lugano, September 16-18, 2009.

23. **Meienberg J**, Neuenschwander S, Patrignani A, Alonso S, Arnold E, Henggeler C, Perez R, Azzarello-Burri S, Steiner B, Spanaus K, Regenass S, Giunta C, Rohrbach M, Carrel T, Steinmann B, Berger W, Matyas G (2009) Large deletion comprising *COL3A1* causes aortic dissection. ESHG 2009, Eur J Hum Genet 17, Suppl. 1:317.

**INVITED ORAL PRESENTATIONS**

1. **Meienberg J** (2015) Precise breakpoint localization of large genomic deletions using PacBio and Illumina next-generation sequencers. Bioinformatics for Third Generation Sequencing Workshop, Institute Pasteur, Lille, France, June 9, 2015.
2. **Meienberg J** (2014) Personalized medicine and orphan diseases. Winter School on Personalized Medicine, University of Zurich, Switzerland, January 22, 2014.
3. **Meienberg J** (2009) Hemizygous deletion comprising *COL3A1*, *COL5A2*, *MSTN*, and *SLC40A1* causes a complex phenotype with aortic dissection. Institute of Medical Genetics, University of Zurich, Switzerland, October 1, 2009.